

Catalogue no. 89-654-X — No. 2014001
ISBN 978-1-100-22832-7

Concepts and Methods

Canadian Survey on Disability, 2012: Concepts and Methods Guide

Social and Aboriginal Statistics Division

February 2014



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca,

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- | | |
|---|----------------|
| • Statistical Information Service | 1-800-263-1136 |
| • National telecommunications device for the hearing impaired | 1-800-363-7629 |
| • Fax line | 1-877-287-4369 |

Depository Services Program

- | | |
|------------------|----------------|
| • Inquiries line | 1-800-635-7943 |
| • Fax line | 1-800-565-7757 |

To access this product

This product, Catalogue no. 89-654-X, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by "Key resource" > "Publications."

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Published by authority of the Minister responsible for
Statistics Canada

© Minister of Industry, 2014

All rights reserved. Use of this publication is governed by the
Statistics Canada Open Licence Agreement (<http://www.statcan.gc.ca/reference/copyright-droit-auteur-eng.htm>).

Cette publication est aussi disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard symbols

The following symbols are used in Statistics Canada publications:

- | | |
|----------------|--|
| .. | not available for any reference period |
| ... | not available for a specific reference period |
| ... | not applicable |
| 0 | true zero or a value rounded to zero |
| 0 ^s | value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded |
| ^p | preliminary |
| ^r | revised |
| x | suppressed to meet the confidentiality requirements of the <i>Statistics Act</i> |
| E | use with caution |
| F | too unreliable to be published |
| * | significantly different from reference category ($p < 0.05$) |

Table of contents

1. Introduction	5
1.1 Survey overview	5
1.2 Background	5
1.3 A Caution to users	6
1.4 Purpose of the Concepts and Methods Guide	6
2. Survey content: concepts and questions	7
2.1 Identifying persons with a disability	7
2.2 Calculating rates of disability in Canada	9
2.3 Measuring the severity of disabilities	9
2.4 Creating a portrait of Canadian adults with disabilities	10
2.5 Questionnaire modules	11
2.6 Linked content from the 2011 National Household Survey	12
2.7 Geographies	13
3. CSD sample design	14
3.1 Target population and coverage	14
3.2 Reference period	15
3.3 Frame drawn from the NHS	15
3.4 Sample design	15
3.5 Estimation domains and stratification	16
3.6 Sample allocation	17
3.7 Sample sizes	18
3.8 The NO sample	18
4. Data collection	20
4.1 Time frame	20
4.2 Communications	20
4.3 Mode of collection	20
4.4 Supervision and quality control	21
4.5 Proxy interviews	21
4.6 Special issues	22
4.7 Response rate	22
5. Data processing	24
5.1 Data capture	24
5.2 Processing steps	24
5.3 Record clean up: in-scope and complete records	25
5.4 Variable recodes and multiple response questions	25
5.5 Flows: response paths, valid skips and question non-response	27

5.6 Coding	27
5.7 Edits and imputations	28
5.8 Derived variables and external NHS-linked variables	29
6. Weighting and creation of final data files	30
6.1 Weighting	30
6.2 File structure and content	36
6.3 Final datasets and data dictionaries	38
6.4 Guidelines for analysis	38
7. Data quality	39
7.1 Overview of data quality evaluation	39
7.2 Sampling errors and bootstrap method	39
7.3 Non-sampling errors	41
8. Differences between the 2012 Canadian Survey on Disability (CSD) and the 2006 Participation and Activity Limitation Survey (PALS)	44
8.1 New method of screening for disability	44
8.2 Other changes to CSD content	45
8.3 NHS filter questions	45
8.4 Change in lag time from the filtering survey	45
8.5 Other methodological changes	46
8.6 Summary and recommendation	46
9. Data dissemination	47
9.1 Data products and services	47
9.2 Reference products	47
9.3 Disclosure control	47
Appendix A – Survey indicators by questionnaire module	48
Appendix B – Disability Severity Score	53
Appendix C – Employment equity questions	58
Appendix D – Extra coding categories and standard classifications used	59
Appendix E – Glossary of survey terms	61
References	69

1. Introduction

1.1 Survey overview

The 2012 Canadian Survey on Disability (CSD) is a national survey of Canadian adults whose everyday activities are limited because of a long-term condition or health-related problem. The CSD was developed by Statistics Canada in collaboration with Employment and Social Development Canada (ESDC) (formerly Human Resources and Skills Development Canada). Input for the survey was obtained from the ESDC Persons with Disabilities Technical Advisory Group which consists of experts in the field of disability, including academics and representatives from various community associations across Canada, as well as members from ESDC and Statistics Canada. The survey was conducted from September 24, 2012 to January 13, 2013.

The CSD is based on a social model of disabilities rather than a medical model. The social model is based on the premise that disability is the result of the interaction between a person's functional limitations and barriers in the environment, including social and physical barriers that make it harder to function day-to-day. Thus, disability is a social disadvantage that an unsupportive environment imposes on top of an individual's impairment (Mackenzie et al, 2009).

The 2012 CSD provides a range of data on different disability types and severities. It measures how often the daily activities of Canadian adults are limited by long-term conditions, health problems and task-based difficulties. The CSD also collects data on use of aids and assistive devices as well as help received or required by respondents. The survey includes questions on the education and employment experiences of persons with disabilities, accommodations made in these areas and ability to get around the community.

Data from the CSD will serve disability and social policy analysts at all levels of government, as well as associations for persons with disabilities and researchers working in the field of disability policy and programs. The CSD will be used to plan and evaluate services, programs and policies for Canadian adults with disabilities to help enable their full participation in society. In particular, information on adults with disabilities is essential for the effective development and operation of the Employment Equity Program. Data on disability are also used to fulfill Canada's international agreement relating to the United Nations Convention on the Rights of Persons with Disabilities.

The 2012 Canadian Survey on Disability (CSD) was based on a sample of persons who reported an activity limitation on the 2011 National Household Survey (NHS) and who were 15 years of age or older as of the date of the NHS, May 10, 2011. Since the NHS excludes the institutionalized population and other collective dwellings, the CSD only covers persons living in private dwellings in Canada. Also, for operational reasons, the population living on First Nation reserves is also excluded. Total sample size for the CSD was approximately 45,500 individuals. The overall response rate was 75%. The CSD provides reliable data on persons with disabilities for each province and territory of Canada. Detailed age breakdowns are also available by province.

1.2 Background

Data on disability in Canada have been collected for over 30 years, reflecting an evolving recognition of the importance of data to support the goal of full participation of persons with disabilities. Concepts and methods used in the production of data on disability have also evolved over time. The first survey on disability in Canada was conducted in 1981, The International Year of the Disabled, shortly after the Canadian Parliamentary Committee on the Disabled and the Handicapped published its report entitled "Obstacles". Among the report's recommendations was that Statistics Canada produce data on persons with disabilities. This launched the Canadian Health and Disability Survey, which was conducted as a supplement to the October 1983 and June 1984 Labour Force Survey. In addition, the 1986 Census included a question about activity limitations that would help to identify persons who were likely to have a disability. Later that year, Statistics Canada used that Census base to select a sample for the Health and Activity Limitation Survey (HALS). The 1986 HALS was conducted to identify Canadians with disabilities and to determine what limitations they experienced and the barriers they faced. A second HALS took place in 1991.

In 1996, no post-censal surveys were conducted. However, in 1998, the federal, provincial and territorial governments released their common disability framework, *In Unison*, calling for the promotion of greater inclusion of persons with disabilities in all aspects of Canadian society. Their 1998 report noted the importance of developing a reliable statistical database on disability and underlined the key role survey data would play in supporting policy development and research in this area.

In 2001, the International Classification of Functioning, Disability and Health (ICF), was approved by all World Health Organization (WHO) member states, including Canada. The ICF defined disability as the relationship between body structures and functions, daily activities and social participation, while recognizing the role of environmental factors. Reflecting this new definition of disability driven by a social model approach, HALS was renamed as the Participation and Activity Limitations Survey (PALS) and was conducted in 2001 and 2006. The new name reflected the fact that the new survey would increase the focus on the participation of persons with activity limitations. As with HALS, PALS was a joint effort by Employment and Social Development Canada and Statistics Canada.

More recently, in 2010, Canada ratified the United Nations Convention on the Rights of Persons with Disabilities. In keeping with Article 31 on Statistics and Data Collection, ESDC launched the New Disability Data Strategy and began developing a new set of questions to identify persons with disabilities, called the Disability Screening Questions (DSQ). The DSQ sought to move more fully towards the social model of disability, to achieve greater consistency in disability identification by type, and to improve coverage of the full range of disability types, especially mental/psychological, learning and memory disabilities. The DSQ were extensively tested qualitatively and quantitatively and were then used for the first time to identify adults with a disability on the 2012 Canadian Survey on Disability (CSD). The CSD also provided a portrait of adult Canadians with disabilities in relation to their participation in society.

1.3 A Caution to users

The concepts and methods used to measure disability in the 2012 Canadian Survey on Disability (CSD) represent a significant departure from those used in the 2006 Participation and Activity Limitation Survey (PALS). The most important change is that the two surveys used a different definition of disability. In the CSD, the new definition was applied by using the new set of disability screening questions (DSQ), reflecting a fuller implementation of the social model of disability. Other significant changes introduced for the CSD (see Chapter 8 below) also affect time series comparability. Because of the major differences in concepts and methods between the 2006 PALS and the 2012 CSD, it is neither possible nor recommended to compare the prevalence of disability over time between these two sources.

1.4 Purpose of the Concepts and Methods Guide

This Concepts and Methods Guide is intended to provide an understanding of the 2012 Canadian Survey on Disability (CSD) with respect to its subject-matter content and its methodological approaches. It is designed to assist CSD data users by serving as a guide to the concepts and questions used in the survey as well as the technical details of survey design, field work and data processing. The guide provides helpful information on how to use and interpret survey results. Its discussion of data quality also allows users to review the strengths and limitations of the data for their particular needs.

Chapter 1 of this guide provides an overview of the 2012 CSD by introducing the survey's background and objectives. Chapter 2 discusses the development and testing of the survey's content, explaining the key concepts and definitions used for the survey. This chapter introduces the CSD questionnaire modules as well as data linkages with the National Household Survey (NHS). Chapters 3 to 6 cover important aspects of the survey methodology, from sampling design, through data collection and processing and ending in the creation of final data files. Chapters 7 and 8 cover issues of data quality and caution users against making comparisons with data from previous Participation and Activity Limitations Surveys (PALS). Chapter 9 outlines the survey products that are more widely available to the public, including data tables, a fact sheet and reference material. Appendices provide more detail on questionnaire indicators, the disability severity score, questions used for assessing employment equity, special coding categories for the survey and standard classifications used. A glossary of survey terms is also provided.

2. Survey content: concepts and questions

2.1 Identifying persons with a disability

In the context of the New Disability Data Strategy of 2010, efforts to better identify persons with disabilities began with the creation of an instrument called the "Disability Screening Questions" (DSQ). The DSQ form the first component of the more comprehensive Canadian Survey on Disability.

The goal in developing the new screening questions was to create a new, comprehensive identification module for persons with disabilities which would:

- move more fully toward the social model of disability,
- identify both type and severity of disability,
- achieve greater consistency in disability identification by type,
- improve coverage of the full range of disability types, in particular, mental/psychological, learning and memory disabilities, and
- be short enough for adoption on a range of general population surveys.

It should be noted that the DSQ were specifically developed for the measurement of disability among adult Canadians but not among children.

Disability screening questions (DSQ)

The social model of disability implies that the presence of a difficulty alone is not sufficient for the identification of a disability - a limitation in daily activities must also be declared. Therefore, the DSQ measure the type and severity of disabilities of Canadian adults by asking questions about how often respondents' daily activities are limited by long-term conditions, health problems and task-based difficulties.

Screeners questions on the DSQ evaluate the presence and severity of 10 distinct types of disabilities related to a health problem or condition that has lasted or is expected to last for six months or more. Screening questions emphasize consistency of measurement across disability types. The questions address the following disability types:

1. Seeing
2. Hearing
3. Mobility
4. Flexibility
5. Dexterity
6. Pain
7. Learning
8. Developmental
9. Mental/psychological
10. Memory

The DSQ also contain a question concerning any other health problem or condition that has lasted or is expected to last for six months or more. This question is associated with the "other" type. However, this type will be counted only if no other limitation has been reported under the 10 types of disability listed above. If there is both a limitation under one of the 10 types and an "other" limitation, the latter will be ignored.¹ An "unknown" type is therefore created for persons who reported only an "other" type of limitation and no other limitation.

¹ A decision was made to ignore the "other" type when there was already a limitation under one of the 10 disability types because it was observed that respondents with a disability that fell under one of the 10 types tended to report the disease that caused their disability under "other". Double counting of disability types was thus avoided.

Qualitative and quantitative testing of disability screening questions

The new set of disability screening questions was drafted following an extensive review of existing disability indicators used in Canada and internationally. Next, the DSQ were tested both qualitatively and quantitatively. Different iterations of qualitative testing were conducted by Statistics Canada's Questionnaire Design Resource Centre (QDRC) in September 2010, March 2011 and May 2011. Qualitative testing was carried out in English and French, in six different locations across Canada. Overall, a total of approximately 125 individual cognitive interviews took place as well as four focus group discussions. Adjustments were made to question wording based on these results.

During the Fall of 2011, the DSQ were also tested quantitatively on two Statistics Canada surveys: the Labour Force Survey and the Canadian Community Health Survey. These tests established the reliability and the validity of the DSQ as an instrument for estimating the prevalence of disability in the adult population within the context of the social model of disability. Following final adjustments to question wording and interviewer instructions, the DSQ were approved for use in the Canadian Survey on Disability (CSD).

It should be noted that an additional shorter version of the DSQ has also been developed as a minimum-content option for identifying persons with a disability on general population surveys at Statistics Canada. This will allow for the reliable identification of persons with disabilities on national surveys of the general population on a more frequent basis, in addition to the more specialized Canadian Survey on Disability (CSD). As a result, more varied data on persons with disabilities in Canada can be produced.

Operational definition of a disability using the DSQ

For each of the 10 disability types, the DSQ always have at least one question on the associated level of difficulty (no difficulty, some difficulty, a lot of difficulty, cannot do) and a question on the frequency of the limitation of daily activities (never, rarely, sometimes, often, always). To meet the definition of a disability for a particular type, the frequency for the corresponding limitation in daily activities must be "sometimes", "often" or "always"), or it must be "rarely" combined with a difficulty level of "a lot" or "cannot do".

Table 2.1 below summarizes the combination of answers to the DSQ that are generally used to identify a disability. This classification applies to the majority of disability types measured on the DSQ.

Table 2.1
Combination of answers to the Disability Screening Questions that were used to identify a disability

How much difficulty do you have...?	How often are your daily activities limited by...?				
	Never	Rarely	Sometimes	Often	Always
No difficulty	No disability	No disability	Disability	Disability	Disability
Some difficulty	No disability	No disability	Disability	Disability	Disability
A lot of difficulty	No disability	Disability	Disability	Disability	Disability
Cannot do	No disability	Disability	Disability	Disability	Disability

Source: Statistics Canada, Canadian Survey on Disability, 2012.

It should be noted that in some situations, these criteria were modified. In particular, a person who reports having a developmental disorder is identified as disabled if the respondent has been diagnosed with this condition, regardless of the level of difficulty or the frequency of the activity limitation reported.

Another noteworthy exception is the "unknown" type, where we do not ask about level of difficulty. A person will have an "unknown" disability only if he or she reports being limited in terms of daily activities "sometimes", "often" or "always" because of this health problem or condition and if he or she reports no other limitation under the 10 disability types.

Lastly, for disabilities involving seeing, hearing, mobility, flexibility and dexterity, which are measured with task-based questions, a response of "no difficulty" does not result in the follow-up question on daily activity limitations. Thus, all "no difficulty" responses for these disability types are classified as "no disability".

2.2 Calculating rates of disability in Canada

One of the primary objectives of the 2012 CSD is to produce rates of disability in Canada. These can be calculated by province and territory, for example, or by age group. Rates of disability are calculated with the following formula:

$$(\text{Persons with a disability} / (\text{Persons with a disability} + \text{Persons without a disability})) \times 100$$

In order to provide such statistics, the methodology of the CSD required not only identifying persons with a disability but also producing estimates of the number of persons without a disability in Canada. Thus, the CSD drew two distinct samples of persons from the 2011 National Household Survey (NHS):

1. Those who were filtered-in by the NHS activity limitation questions (called the YES sample) and who would proceed through the Disability Screening Questions (DSQ) of the Canadian Disability Survey (CSD) for the more precise identification of persons with a disability, and
2. Those who were filtered-out by the NHS activity limitation questions (called the NO sample) and who were automatically considered persons without a disability.

Details about these methods are provided in Chapter 3 of this guide.

2.3 Measuring the severity of disabilities

Usefulness of a severity score

It is clear from previous research using the 2006 Participation and Activity Limitation Survey that disability severity is a strong predictor of the reduced participation of people with disabilities in several domains of everyday life (Federal Disability Report, 2010). People with severe disabilities are less likely than their counterparts with milder disabilities to participate in the labour force, to attend and complete post-secondary education programs, and to participate in community activities. Those with severe disabilities are also more likely to be in need of supports and services, such as aids, devices, caregiving, medical specialists and income supports.

The inclusion of disability severity is an important consideration in analyses of the participation of people with disabilities. The ready-to-use, consistent disability severity score included in the CSD file enables analysts to develop more accurate inferences about the situation currently faced by persons with disabilities.

The Severity Score

A severity score was developed using the Disability Screening Questions (DSQ). For each of the 10 disability types, a score is assigned using a scoring grid that takes into account both the frequency of the activity limitations (never, rarely, sometimes, often, or always) and the intensity of the difficulties (no difficulty, some difficulty, a lot of difficulty, or cannot do). The score increases with the frequency of the limitation and the level of difficulty.

A global severity score is derived based on all disability types. A person's global severity score is calculated by taking the average of the scores for the 10 disability types. Consequently, the more types of disability a person has, the higher his or her score will be.

Overall, the global score meets the following three criteria:

- it increases with the number of disability types;
- it increases with the level of difficulty associated with the disability;
- it increases with the frequency of the activity limitation.

To make the severity score easier to use, severity classes were established. It is important to understand that the name assigned to each class is simply intended to facilitate its use. It is not a label or judgement concerning the person's level of disability. In other words, the classes should be interpreted as follows: people in class 1 have a less severe disability than people in class 2; the latter have a less severe disability than people in class 3; and so on:

- 1 = mild disability
- 2 = moderate disability
- 3 = severe disability
- 4 = very severe disability

The breakdown of persons with a disability across the four severity classes is shown in the table below.

Table 2.2
Distribution of persons with a disability, by level of severity, Canada

Severity class	Persons with a disability	
	number	%
Class 1 = mild	1,195,590	31.7
Class 2 = moderate	747,980	19.8
Class 3 = severe	849,540	22.5
Class 4 = very severe	982,810	26.0

Source: Statistics Canada, Canadian Survey on Disability, 2012.

For additional information on the methods used to derive severity scores, see Appendix B.

2.4 Creating a portrait of Canadian adults with disabilities

In addition to the Disability Screening Questions (DSQ), detailed content was identified for the Canadian Survey on Disability (CSD) which provided a picture of the issues affecting Canadian adults who have difficulty carrying out their daily activities. The CSD covered the use of aids and assistive devices as well as help received or required by persons with disabilities. Data priorities also included the education and employment experiences of adults with disabilities, accommodations made in these areas, the ability to get around the community using public transportation, and sources of income.

Survey indicators in these content areas were drawn from a number of sources, including the 2006 Participation and Activity Limitation Survey (PALS). Relevant standardized and well-established measures used on other Statistics Canada surveys were also gathered and reviewed as potential indicators. Those indicators considered most important were then modified for increased relevance to the population of persons with disabilities.

A draft questionnaire was tested in March 2012. Qualitative testing involved 50 cognitive interviews in both English and French, in three locations across Canada. Results of testing led to recommendations towards a final set of robust questions for the full Canadian Survey on Disability. Improvements to the wording of questions, interviewer instructions and the flow of the questionnaire were implemented.

Questions in the 2012 CSD were designed for use in a Computer Assisted Interviewing (CAI) environment. CAI enables more complex question flows to be built into the questionnaire and includes many features to maximize data quality, such as consistency verifications across questions. Specifically for the CSD, the questionnaire consisted of a Computer Assisted Telephone Interview (CATI). The new CATI questionnaire application underwent extensive modular and end-to-end testing.

One final step was taken with respect to content development for the 2012 CSD: linkage to data from the 2011 National Household Survey (NHS). Since the 2012 CSD drew its sample from the 2011 NHS (see Chapter 3 for details), relevant information from the NHS can be combined with information provided during the 2012 CSD interview. This approach reduces the number of questions that need to be asked on the CSD and provides for a richer portrait of persons with disabilities for CSD data users. In addition, since a sample of persons without a disability was drawn from the NHS for the purpose of calculating rates of disability in Canada (see Chapter 3 for more detail), NHS data linkage also provide a rich source of data on people without disabilities, allowing important comparisons of characteristics to be made between the two groups.

More details of this process of NHS data linkage are provided in Section 2.6 below.

A special note about age data

Age is a core demographic factor of interest in the analysis of disability in Canada. When using age as a component of research with 2012 CSD data, or in combination with linked data from the 2011 NHS, it is important for users to keep in mind the different reference periods involved. Section 6.2 of this guide provides an explanation of these survey reference periods. With respect to age, it is important to note that data collected from respondents in the context of the NHS were collected on May 10, 2011 while data from the CSD were collected 16 to 20 months later, between September 2012 and January 2013. So, for example, CSD respondents who were 15 years of age at the time of the NHS were 16 to 20 months older at the time of the CSD. With respect to particular research studies that may be sensitive to this time lag, data users will have the option of selecting an age indicator based on the 2011 NHS reference period or an age variable based on the date of the 2012 CSD interview. Section 6.2 provides an understanding of the use of survey reference periods in relation to different types of data analyses that may be of interest to users.

2.5 Questionnaire modules

Listed below are the questionnaire modules found on the 2012 Canadian Survey on Disability. Each are described in detail in Appendix A.

The Canadian Survey on Disability – 2012

Introduction

- Purpose of the survey
- Voluntary nature of survey
- Explanation of data linkage with NHS

Disability screening questions

- Filter questions
- Screeners

Disability questions

- Main underlying condition
- Aids and assistive devices
- Recurring aids and assistive devices questions
- Medication use
- Received and needed help with everyday activities

Education

- Recent school attendance
- Past school attendance
- Educational experiences
- Educational background

Employment

- Labour force status
- Employment details
- The unemployed
- Not in labour force
- Retirement
- Workplace training
- Employment modifications
- Labour force discrimination

Getting around the community

Source of income

In addition to information on the questionnaire modules in Appendix A, Appendix D lists the extra question categories created during survey coding as well as the standard classifications used to create indicators for open-ended survey questions. The Canadian Survey on Disability questionnaire can be found on the Statistics Canada website.

A comprehensive description of all the variables available from the survey data is provided in the 2012 CSD Data Dictionaries (codebooks). For details on how to obtain the data dictionaries, contact Statistics Canada Client Services.

2.6 Linked content from the 2011 National Household Survey

The Canadian Survey on Disability (CSD) draws its sample from among 2011 National Household Survey (NHS) respondents (see Chapter 3 for details). At the outset of the 2012 CSD interview, all respondents were told about the plans to link the CSD survey data with the information that they provided on the NHS. All linked information is kept confidential and used for statistical purposes only.

The specific benefits of a CSD-NHS record linkage are reduced response burden for the target population of the CSD, the establishment of survey weights which are crucial to providing valid estimates, and the creation of a comprehensive microdata file on persons with disabilities in Canada. Together, data from these two sources provide a detailed statistical portrait of persons with disabilities in Canada - data which are not available from any other source.

As explained in more detail in Chapter 3, the CSD also drew a sample of persons without a disability from the NHS. NHS linkage thus allows data users to compare characteristics of persons with a disability and persons without a disability.

Approximately 200 NHS variables were linked to the final CSD file for 2012, both for persons with a disability and for persons without a disability. The list below shows the type of NHS variables that have been appended to the CSD analytical files.

- Household level variables
 - Geography, including census metropolitan areas
 - Housing, including tenure, number of rooms in dwelling and need for repairs
 - Family, including presence of spouse/partner in household, number of children in census family
- Person-level variables
 - Activity limitations, including activity difficulties/reductions at home, work and school
 - Aboriginal identity and ancestry
 - Demographics, including age, marital status and common-law status, place of birth
 - Immigration, citizenship and ethnicity
 - Education, including school attendance, location of study, postsecondary certificates, diplomas and degrees, types obtained, major field of study
 - Employment, including labour force status, weeks worked in 2010, class of worker, full-time or part-time work
 - Place of work, including place of work status, type of commute and distance
 - Mobility, including mobility status 1 and 5 years ago
 - Income, including family income, employment income, low-income status (before and after tax)
 - Language, including knowledge of official languages, mother tongue, language spoken at home and language of work

It is important to note that these NHS variables refer to each respondent's situation on the day of the 2011 NHS, that is, as of May 10, 2011. Thus, for 2012 CSD respondents, users should be aware that in some cases, the respondent may have moved, had a change in the composition of their household, or had a change in employment between the date of the 2011 NHS and the date of the 2012 CSD interview. In other words, some of the information provided by the NHS may not be reflective of the respondent's situation when the CSD interview took place.

A complete list of the NHS variables and their specifications are provided in the 2012 CSD Data Dictionaries.

2.7 Geographies

The 2012 Canadian Survey on Disability (CSD) was designed to produce reliable data for each of the provinces and territories. Other geographic variables are also available in the 2012 CSD database, based on geographies from the National Household Survey, such as Census Metropolitan Areas (CMAs). In addition, geographies will include health regions across Canada which represent administrative areas or regions as used by health authorities. However, users should note that not all CSD survey data can be cross-tabulated or analyzed at these more detailed levels of geography. Some data tables will be possible but the reliability of data estimates at these levels of geography will need to be examined on a case by case basis.

The NHS Dictionary defines geographies relevant to the CSD.

More details on health regions can be found on the Statistics Canada website.

3. CSD sample design

3.1 Target population and coverage

The **population covered** by the Canadian Survey on Disability (CSD) consists of all persons aged 15 and over (on Census/NHS Day, May 10, 2011) who have an activity limitation or a participation restriction associated with a physical or mental condition or health problem and were living in Canada at the time of the Census/NHS. This includes persons living in private dwellings in the 10 provinces and three territories. However, for operational reasons, the population of Indian reserves is excluded.

The CSD's sample is based on responses to questions in the 2011 National Household Survey (NHS). All persons who answered yes to at least one of the NHS filter questions on activity limitations (see Figure 1) were included in the CSD frame. These persons are members of what is referred to as the **YES population**. Within this group, people whose daily activities are limited due to a long-term condition or health problem are members of the **target population** of persons with a disability.

Although the CSD does not cover the population that answered no to the NHS filter questions (referred to as the **NO population**), a sample of these persons is included in the CSD analytical file (see Section 3.8). These persons are all considered persons without a disability. There are also a number of persons without a disability in the YES population; they did not report any activity limitations in the CSD's DSQ module. As we will see later, the sample of persons without a disability is used in two ways: to calculate disability rates and to compare the NHS characteristics of persons with a disability and persons without a disability.

Figure 3.1
2011 NHS filter questions on activity limitations

ACTIVITIES OF DAILY LIVING

National Household Survey – N1 – question 7

Does this person have any **difficulty** hearing, seeing, communicating, walking, climbing stairs, bending, learning or doing any similar activities?

Yes, sometimes
Yes, often
No

National Household Survey – N1 – question 8

Does a physical condition or mental condition or health problem **reduce the amount or kind of activity** this person can do:

(a) at home?

Yes, sometimes
Yes, often
No

(b) at work or at school?

Yes, sometimes
Yes, often
No
Not applicable

(c) in other activities, for example, transportation or leisure?

Yes, sometimes
Yes, often
No

Source: Statistics Canada, 2011 National Household Survey.

3.2 Reference period

The Canadian Survey on Disability (CSD) represents the population aged 15 and over on Census/NHS Day, **May 10, 2011**. However, because of the time lag between the Census/NHS and CSD collection, all the information collected in the CSD represents the population's characteristics as measured in the **fall of 2012**. To understand how these two reference periods affect the use and interpretation of CSD data, refer to section 6.2.

3.3 Frame drawn from the NHS

The sampling frame was constructed from the Census and NHS response database (RDB). The database contained all responses received via the various reporting modes (Internet, paper questionnaires, personal interviews, etc.).

The frame had to go through a number of processing steps to ensure that all the information would be as complete as possible:

- when the age was missing for a person on the NHS questionnaire, it had to be imputed on the basis of responses to NHS questions (for example, data on education, employment and income);
- various data sources were used to detect errors in telephone numbers or find missing telephone numbers or addresses;
- preliminary weights for the NHS had to be derived so that estimates of the population size in each stratum could be computed for sample allocation.

3.4 Sample design

The CSD's sample design can be viewed as a three-phase design, where the first two phases are for the selection of the NHS sample itself and the third phase is for the selection of the CSD sub-sample.

Phase 1

The initial phase was the sample selection of the NHS itself. The NHS questionnaire comes in two main versions, the N1 form and the N2 form.

The N1 form is completed by self-enumeration and is administered to approximately one in three households in most parts of Canada (N1 regions). Other than the basic census demographic questions (name, sex, date of birth, legal marital status, common-law status, relationship to person 1, various language questions and the consent question to release the data in 92 years), the NHS N1 form includes questions on labour market activity, income, education, activity limitations, citizenship, housing, ethnic origin, and so on.

The N2 form, identical in content to the N1 form, except for some adapted examples and questions excluded², is administered by personal interview to all households in remote areas and Indian reserves (N2 regions).

In sampling terminology, the NHS sample design is a stratified systematic sample of occupied private dwellings with a constant sampling rate of 1/3 in N1 regions, and with complete enumeration in N2 regions.

Phase 2

Non-response to the NHS was not inconsequential and a new strategy was put in place to mitigate the potential effect of non-response bias. Hence a subsample of non-respondent households was selected for non-response follow-up (NRFU). Consequently, this subsample carries a greater sampling weight to reflect both probabilities of selection at each phase. The subsample was selected in such a way as to ensure good representation of collection units (CUs) with higher concentrations of groups at risk of not responding³. It is

² The questions on citizenship (question 10), immigrant status received (question 11) and year of immigration (question 12) are not asked of people living on Indian reserves and settlements who were enumerated using the N2 form.

³ Many simulations were carried out in preparation for NHS collection in an effort to identify the groups with the greatest risk of not responding to this survey. CUs with high concentrations of these target groups were oversampled for non-response follow-up.

important to note that non-response follow-up was not performed in canvasser areas (remote regions and Indian reserves and Indian settlements, also known as N2 regions). This NRFU subsample represents about 1 out of 3 non-responding households. The design in this second phase is a model assisted random sample.

Phase 3

In the third phase, the CSD sample was selected from the group of individuals who responded to the NHS (including the NRFU subsample) and reported activity limitations to the NHS. Excluded from selection were persons living on Indian reserves, as well as persons less than 15 years old on May 10, 2011.

The CSD sample was selected so that there would be a sufficiently large sample in each estimation domain, as explained in section 3.5.

3.5 Estimation domains and stratification

Domains of estimation are groups of units for which estimates are targeted with an "acceptable" level of precision. These domains of estimation for the CSD consist of the provinces cross-classified with the following age groups:

- 15 to 24
- 25 to 44
- 45 to 64
- 65 to 74
- 75 and over

For each of the three territories, the estimation domain includes a single age group (15+). For Prince Edward Island, the first two age groups had to be combined because of their very small population sizes.

Each estimation domain was further sub-divided into strata to take the severity⁴ of the activity limitation and the NHS sample design into account. To control for severity, those people who answered "yes, often" at least once on the NHS filter questions were placed in the "Often" stratum, while those who answered "yes, sometimes" at least once but never "yes, often" were placed in the "Sometimes" stratum. To control for the NHS sample design, two characteristics were considered: being among the NHS's initial respondents (the IR strata) as opposed to the non-response follow-up respondents (the NRFU strata), and living in a remote region (the N2 strata) or not (the N1 strata). Note that no non-response follow-up was carried out in remote regions. Controlling for these characteristics effectively grouped people with similar preliminary weights.

Hence, each estimation domain was divided into six possible strata defined as follows:

- N1 region - initial respondent - "often"
- N1 region - initial respondent - "sometimes"
- N1 region - NRFU respondent - "often"
- N1 region - NRFU respondent - "sometimes"
- N2 region - initial respondent - "often"
- N2 region - initial respondent - "sometimes"

Note that not all 6 strata necessarily occurred in each estimation domain. All persons meeting the conditions for the CSD frame were then classified into these estimation domains and strata prior to sample selection.

⁴ The concept of "severity" here is not the same used for the severity score, which was described in section 2.3. Here, we only take into account the frequency of limitation reported to the NHS filter questions.

3.6 Sample allocation

The sample sizes were determined in such a way that, for each estimation domain, one could estimate a minimum proportion with a maximum coefficient of variation (CV) of 16.5%. At Statistics Canada, 16.5% is often used as the upper limit for the CV of an acceptable estimate. The minimum proportion to estimate in each estimation domain is shown in the table below. A design effect of 1.2 was assumed for these calculations. In other words, it was assumed that in the estimation domains, the variance that would be obtained with the CSD's sample design would be 20% higher than the variance that would be obtained if a simple random sample of the same size in each domain was selected.

Table 3.1

Minimum proportions to estimate with a maximum coefficient of variation (CV) of 16.5%, by estimation domain

Province/Territory	Age groups				
	15 to 24	25 to 44	45 to 64	65 to 74	75 and over
	%				
Newfoundland and Labrador	9.0	8.0	8.0	11.0	11.0
Prince Edward Island	9.0		8.0	11.0	11.0
Nova Scotia	9.0	8.0	8.0	11.0	11.0
New Brunswick	9.0	8.0	8.0	11.0	11.0
Quebec	9.0	8.0	8.0	11.0	11.0
Ontario	9.0	8.0	8.0	11.0	11.0
Manitoba	9.0	8.0	8.0	11.0	11.0
Saskatchewan	9.0	8.0	8.0	11.0	11.0
Alberta	9.0	8.0	8.0	11.0	11.0
British Columbia	9.0	8.0	8.0	11.0	11.0
Yukon			8.0		
Northwest Territories			8.0		
Nunavut			8.0		

Source: Statistics Canada, Canadian Survey on Disability, 2012.

A method of optimal allocation among the strata in a particular domain was used, taking into account the expected non-response and the expected false positive rate (persons who reported activity limitations in the NHS but have no disability according to the DSQ) in each stratum. This allocation depended in part on the NHS weights adjusted for non-response. It should be noted that at the time of allocation, those weights had not yet been calculated. Consequently, preliminary weights were calculated solely for the purposes of the allocation. However, the final NHS weights were used in the CSD weighting process. For background information on NHS weighting, see Morel and Nambeu (2013).⁵

Following sample allocation, it sometimes happened that more units were needed in one stratum than the number available in the sampling frame. In that case, all the units in the stratum had to be selected (making it a take-all stratum) and then the number of units needed to make up the shortfall was selected from the other strata in the same estimation domain in proportion to their allocation to those strata.

5 Morel, J., and Nambeu, C. (2013). "National Household Survey: Weighting and estimation update". Presented at the SSC conference in May 2013.

3.7 Sample sizes

The CSD's final sample size (the YES sample) for each province and territory is shown in Table 3.2 below.

Table 3.2
Sample sizes by province and territory

Province/Territory	Sample sent to collection
Newfoundland and Labrador	3,792
Prince Edward Island	2,809
Nova Scotia	3,769
New Brunswick	4,146
Quebec	4,254
Ontario	4,563
Manitoba	5,081
Saskatchewan	4,372
Alberta	4,740
British Columbia	5,197
Yukon	913
Northwest Territories	836
Nunavut	971
TOTAL	45,443

Source: Statistics Canada, Canadian Survey on Disability, 2012.

3.8 The NO sample

As previously mentioned, the CSD analytical file is used in part to estimate disability rates for various geographies and to compare the characteristics of persons with a disability and persons without a disability. For that purpose, the analytical file must also include a representative sample of persons without a disability, which is not the case when only the YES sample is considered. A sample from the NO population therefore had to be selected, i.e. persons who reported no activity limitations in the NHS filter questions. The underlying assumption here is that the members of the NHS's NO population are less likely to have a disability, or that if they have a disability, it is very mild. The decision not to cover the NO population in the CSD also has to do with the cost involved, since a very large sample of that population is needed to find persons with a disability, which is very expensive during collection.

The NO sample was not sent to the field, since each individual is considered not to have a disability. The analytical file contains a large number of characteristics from the NHS for this sample and for the YES sample, which makes it possible to compare persons with a disability and persons without a disability. With this NO sample, analysts are also able to produce the denominators required to calculate the disability rates for a number of subgroups of the Canadian population. To keep the analytical file from getting too large, a sample of the NO population was selected instead of taking the entire population. The size of this sample was set so that the same precision as is required for the YES sample would be achieved (see Table 3.1), with a maximum CV of 7%.⁶ In selecting the sample, the NO population was stratified by province, sex and five-year age group, with the oldest group being 75 and over (thus, it was more detailed than the YES sample). The NHS design was also taken into account for the NO sample, where strata were further divided into N1 vs N2 regions and initial vs NRFU respondents. The sample sizes for the provinces are shown in the table below.

⁶ For Prince Edward Island, we set the CV at 10% for the 65-to-74 age group and 15% for the 75-and-over age group to avoid selecting all available units in the sampling frame.

Table 3.3
NO sample sizes by province and territory

Province/Territory	NO sample (not sent to collection)
Newfoundland and Labrador	11,813
Prince Edward Island	6,643
Nova Scotia	12,062
New Brunswick	11,997
Quebec	12,455
Ontario	12,486
Manitoba	12,152
Saskatchewan	12,103
Alberta	12,338
British Columbia	12,426
Yukon	2,499
Northwest Territories	2,578
Nunavut	2,446
TOTAL	123,998

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Since all units in the NO population are considered not to have a disability, the severity score for each unit in the file was set to 0, and the severity class was set to 0.

4. Data collection

4.1 Time frame

The Canadian Survey on Disability (CSD) was conducted from September 24, 2012 to January 13, 2013. As a post-censal / post-NHS survey, it followed the 2011 National Household Survey which was conducted on May 10, 2011. Thus, a time lag of 16 to 20 months existed between the two surveys.

4.2 Communications

In the months leading up to data collection for the 2012 Canadian Survey on Disability (CSD), promotional activities took place to raise awareness of the survey and to encourage participation.

A webpage appeared on the Statistics Canada website which included "Questions and Answers" for respondents, a survey description, background information on the survey and its methodology, and a link to the questionnaire. A special website icon was designed by Statistics Canada and distributed to the ESDC Persons with Disabilities Technical Advisory Group (TAG), and they helped to promote the survey by displaying the survey icon on their organizations' websites, thus providing a link to Statistics Canada's webpage for the CSD.

Prior to data collection, an introductory letter and survey brochure were sent to respondents which informed them about the upcoming survey and impressed upon them the importance of their participation. The letter was provided in both official languages. For survey collection in Nunavut, the letter was also provided in Inuktitut (Baffin dialect). In addition, a Braille insert was included with the introductory letters which provided a contact telephone number so that those with visual impairments could contact Statistics Canada for more information.

4.3 Mode of collection

The questions in the 2012 Canadian Survey on Disability (CSD) were administered using Computer Assisted Telephone Interviews (CATI). Accommodations were made to maximize direct participation of persons with disabilities by providing all respondents with an email address, a TTY telephone number (Telecommunications Device for the Hearing Impaired) and a contact telephone number written on a Braille insert that accompanied the introductory letter. Where alternate arrangements were requested, personal interviews were conducted using a paper and pencil interview (PAPI) version of the questionnaire.

A PAPI questionnaire was also used to conduct in-person interviews with some respondents in the Northwest Territories who did not have a telephone and could not otherwise have been contacted. It was known prior to the survey that a large proportion of dwellings in the western territories did not have telephones and this meant that coverage would be generally limited to NHS respondents who had a telephone. Prior to the start of collection, Statistics Canada was able to establish a partnership with the NWT Bureau of Statistics to hire a team of interviewers to help conduct personal interviews. As a result, a sample of dwellings which did not have telephones was selected and a PAPI version of the questionnaire was used for house-to-house data collection.

Across Canada, respondents were interviewed in the official language of their choice – English or French. The questionnaire was also translated into Inuktitut and interviews in that language were made available for respondents in Nunavut.

The time required to complete the survey varied from person to person but, on average, the survey took approximately 40 minutes to complete.

4.4 Supervision and quality control

All Statistics Canada interviewers were under the supervision of senior interviewers who were responsible for ensuring that interviewers were familiar with the concepts and procedures of the surveys to which they were assigned. Senior interviewers were also responsible for periodically monitoring the interviewers to ensure that standard procedures were being followed.

Interviewers were trained on the survey content and the computer-assisted interviewing application. In addition to classroom presentations and a period of self-study, the interviewers completed a series of mock interviews to become familiar with the survey and its concepts and definitions.

4.5 Proxy interviews

Since disability is difficult to measure and very subjective, interviewers were asked to make every effort to conduct the interview with the selected person. However, in the following circumstances, a proxy interview was acceptable:

- The selected person was away for the duration of the survey.
- The selected person spoke neither English nor French.
- The selected person was unable to participate because of mental or physical health problems.
- A parent insisted on responding for his or her child aged 15 to 17.

In order to be accepted as a proxy respondent, the person responding must be:

- an adult who speaks either English or French;
- reachable during the survey's data collection period;
- the person most knowledgeable, or among the most knowledgeable, about the selected person's difficulties and challenges related to activity limitations and participation restrictions.

A total of 5,164 proxy interviews were conducted, and 93% of them were completed.

The table below shows the distribution of proxy interviews by respondent's age and reason for the proxy interview (excluding 320 persons for whom the reason is unknown).

Table 4.1
Distribution of the number of proxy interviews by respondent's age and reason for proxy interview

Age group	Reason for doing a proxy interview				TOTAL
	Health	Absent	Language	Parent* insists on answering	
15 to 24	835	431	30	328	1,624
25 to 44	360	150	100	8	618
45 to 64	235	132	167	7	541
65 to 74	335	62	178	5	580
75 and over	863	47	216	7	1,133
TOTAL	2,628	822	691	355	4,496

*In the case of young respondents, the proxy is typically a parent or guardian, but for older respondents, it may have been the person's spouse or another family member.

Source: Statistics Canada, Canadian Survey on Disability, 2012.

4.6 Special issues

Interviewers were instructed to make all reasonable attempts to obtain a completed interview with the selected member of the household. Those who refused to participate were re-contacted up to two more times to explain the importance of the survey and to encourage their participation. For cases in which the timing of the interviewer's call was inconvenient, an appointment was arranged to call back at a more convenient time. For cases in which there was no one home, numerous call backs were made.

Special issues arose in relation to data collection for the Canadian Survey on Disability (CSD) which were addressed with extra coordination in the field and corrective adjustments to survey methods. For instance, due to the extended time lag between the 2011 NHS and the 2012 CSD, the contact information for respondents was sometimes out of date. This affected the successful delivery of introductory letters for respondents who were no longer at the address provided at the time of the NHS. In terms of out-of-date telephone numbers, interviewers were provided with special tracing instructions on ways to locate these respondents in order to complete the interview.

As occasionally happens, the collection of the 2012 CSD occurred during the same time period as several other surveys conducted by Statistics Canada, creating a potential risk for respondent-burden. Careful planning and adjustments to survey design were implemented to address and minimize the burden to households for survey respondents.

A comprehensive description of adjustments applied to survey weights as a result of special issues arising during data collection are provided in Section 6.1.

4.7 Response rate

Collection for the Canadian Survey on Disability (CSD) ended with a response rate of 74.6%. This response rate is the number of complete respondents (with or without a disability) divided by the number of cases sent to collection minus the out-of-scope cases. Out of scope cases include people who have died, who were institutionalized, have emigrated, moved to an Indian reserve, are members of the Canadian Forces, are visitors to Canada (misclassified during the NHS) or have an invalid age. Hence, the response rate reflects the percentage of cases that completed the interview relative to the number of cases that should have completed it (which is why the out-of-scope cases are excluded from the denominator).

$$\text{Response rate} = \text{Completed cases} / (\text{Cases sent to collection} - \text{Out-of-scope cases})$$

The tables below show the response rates by province and territory, and by age group.

Table 4.2
Response rates by province and territory

Province/Territory	Sent to collection	Out of scope	Completed	Response rate
	number			%
Newfoundland and Labrador	3,792	156	2,733	75.2
Prince Edward Island	2,809	118	2,126	79.0
Nova Scotia	3,769	141	2,819	77.7
New Brunswick	4,146	109	2,761	68.4
Quebec	4,254	134	3,182	77.2
Ontario	4,563	117	3,243	72.9
Manitoba	5,081	158	3,645	74.0
Saskatchewan	4,372	154	3,356	79.6
Alberta	4,740	119	3,522	76.2
British Columbia	5,197	122	3,635	71.6
Yukon	913	20	655	73.3
Northwest Territories	836	35	545	68.0
Nunavut	971	37	600	64.2
Total	45,443	1,420	32,822	74.6

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Table 4.3
Response rates by age group

Age groups	Sent to collection	Out of scope	Completed	Response rate
	number			%
15 to 24	9,944	121	6,633	67.5
25 to 44	11,672	119	7,831	67.8
45 to 64	10,487	180	8,182	79.4
65 to 74	6,481	244	5,282	84.7
75 and over	6,859	756	4,894	80.2

Source: Statistics Canada, Canadian Survey on Disability, 2012.

5. Data processing

5.1 Data capture

Responses to survey questions were captured directly by the interviewer using a computerized questionnaire. This involved an application developed using Blaise software in a Computer Assisted Telephone Interview (CATI) system. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission. Data from the CSD paper questionnaires were also entered into the CATI system once forms were returned from the field and were thus electronically captured for further processing.

Some editing of data is done directly at the time of the interview. Specifically, where a particular response appears to be inconsistent with previous answers or outside of expected values, the interviewer is prompted, through message screens on the computer, to confirm answers with the respondent, and, if needed, to modify the information. The response data are subjected to further edit and imputation processes once they arrive in head office.

5.2 Processing steps

Data processing involves a series of steps to convert the questionnaire responses from their initial raw format to a high-quality, user-friendly database involving a comprehensive set of variables for analysis. A series of data operations are executed to clean files of inadvertent errors, edit the data for consistency, code open-ended questions, create useful variables for data analysis, and finally to systematize and document the variables for ease of analytical usage.

The 2012 CSD used a new set of social survey processing tools developed at Statistics Canada called the "Social Survey Processing Environment" (SSPE). The SSPE involves SAS software programs, custom applications and manual processes for performing the following systematic processing steps:

- Receipt of raw data
- Clean up
- Recodes
- Flows
- Coding
- Edits and imputations
- Derived variables
- Creation of final processing file
- Creation of dissemination files

5.3 Record clean up: in-scope and complete records

Following the receipt of raw data from the electronic questionnaire applications, a number of preliminary cleaning procedures were implemented for the 2012 CSD at the level of individual records. These included the removal of all personal identifier information from the files, such as names and addresses, as part of a rigorous set of ongoing mechanisms for protecting the confidentiality of respondents. Duplicate records were resolved at this stage. Also part of clean up procedures was the review of all respondent records to ensure each respondent was "in-scope" and had a sufficiently completed questionnaire. Specific criteria for respondents are outlined below.

1. To be "in scope" for the 2012 CSD, respondents must be at least 15 years of age as of the day of the National Household Survey (NHS), May 10, 2011. In-scope respondents include two groups: 1) those who were screened-in upon completing the Disability Screening Questions (DSQ) and were therefore part of the disability population and 2) those who were screened-out by the DSQ and were thus considered non-disabled. Both groups remain in the survey database.
2. To have a "complete" questionnaire, respondents who met the criteria of the disabled population must have completed the minimum number of critical questions on the survey: those required to produce data tables for persons with disabilities as required by the 1995 Employment Equity Act. These questions are listed in Appendix C.
3. To have a "complete" questionnaire, respondents who were assigned to the non-disabled population must only have completed the DSQ. The rest of the survey questions did not apply to them.

Those that did not meet these criteria were removed from the database, and classified as non-respondents.

5.4 Variable recodes and multiple response questions

This stage of processing involved changes at the level of individual variables. Variables could be dropped, recoded, re-sized or left as is. Formatting changes were intended to facilitate processing as well as analysis of the data by end-users. One such change is the dropping of the letter Q which appeared on every variable name on the questionnaire at the beginning of each question number (shown below).

Another change at the variable level was the conversion of multiple-response questions ("Mark-all-that-apply" questions) to corresponding sets of single-response variables which are easier to use. For each response category associated with the original question, a variable was created with "yes/no" response values. An example is provided below. This process is called "destringing" the variables and allows easier statistical tabulation of the variables by end-users.

Original multiple-response question:

DRV_Q06 What are the reasons you have difficulty using public transit or specialized transit service?

INTERVIEWER: Examples of specialized transit services: Handi-Transit, Wheel-Trans, Para-Transpo, etc.

Mark all that apply. Read categories to respondent

- 01 Service is not available when you need it
- 02 Booking rules don't allow for last minute arrangements
- 03 Difficulty getting to or locating bus stops
- 04 Difficulty getting on or off the vehicle
- 05 Difficulty seeing signs or notices, stops or hearing announcements
- 06 Overcrowding
- 07 Difficulty requesting service
- 08 Difficulty interpreting schedules
- 09 Difficulty transferring or completing complicated transfers
- 10 Your condition or health problem is aggravated when you go out
- 11 Too expensive
- 12 Other reason
- DK, RF

Final variables in single-response "yes/no" format:

DRV_06A What are the reasons you have difficulty using public transit or specialized transit service?

- Service is not available when you need it

1 Yes
2 No
DK, RF

DRV_06B What are the reasons you have difficulty using public transit or specialized transit service?

- Booking rules don't allow for last minute arrangements

1 Yes
2 No
DK, RF

DRV_06C What are the reasons you have difficulty using public transit or specialized transit service?

- Difficulty getting to or locating bus stops

1 Yes
2 No
DK, RF

...additional "yes/no" questions for each response category, as indicated, from - Difficulty getting on or off the vehicle to - Too expensive... and including the last category:

DRV_06L What are the reasons you have difficulty using public transit or specialized transit service?

- Other reason

1 Yes
2 No
DK, RF

In some cases, multiple-response questions on the survey have corresponding follow-up questions which are also multiple-response items. Thus, a single question on the questionnaire can become a large array of variables on the final database.

For example, **AAD_Q10** asks: **Which aids or assistive devices do you need but do not have?** A total of 17 response categories of assistive devices are presented (and one additional category was created in coding – see discussion below – for a total of 18 final categories). Thus, **AAD_Q10** will become 18 separate variables in a "yes/no" format: **AAD_10A** through **AAD_10R**. This question is then followed up with **AAD_Q11 - Why do you not have (refers to AAD_Q10 response(s))?** For this follow-up question, a total of 8 response categories of reasons are presented. Thus, for **AAD_Q11**, a total array of 144 (18 x 8) "yes/no" variables are created, where the 18 devices correspond to letters A through R, and the 8 reasons correspond to letters A through H: **AAD_11AA**, **AAD_11AB**, **AAD_11AC**... **AAD_11AH** through **AAD_11RA**, **AAD_11RB**, **AAD_11RC**... **AAD_11RH**.

5.5 Flows: response paths, valid skips and question non-response

Another set of data processing procedures for the 2012 CSD was the verification of questionnaire flows or skip patterns. All response paths and skip patterns that were built into the questionnaire were verified to ensure that the universe or target population for each question was accurately captured during processing. Special attention was paid to distinctions between valid skips and non-response, an important distinction for statistical analysis. These concepts are explained below in order to assist users to better understand question universes as well as statistical outputs for CSD survey variables.

Response – an answer directly relevant to the content of the question that can be categorized into pre-existing answer categories, including 'other-specify'.

Valid skip – indicates that the question was skipped because it did not apply to the respondent's situation, as determined by valid answers to a previous question. In such cases, the respondent is not considered to be part of the target population or universe for that question. As noted below, where a question was skipped due to an undetermined path (that is, a "don't know" or "refusal" to a previous question caused the skip), the respondent is coded to "Not stated" for that question.

Don't know – the respondent was unable to provide a response for a number of reasons - for example, due to difficulty remembering or because they were responding for someone else.

Refusal – the respondent preferred not to respond, perhaps due to sensitivity of the question.

Not stated – this indicates that the question response is missing and there is an undetermined path for the respondent, such as when a respondent did not answer the previous filter question or where an inconsistency was found in a series of responses.

Suppressed - For NHS variables only, an additional type of non-response is the suppressed value as a result of consistency edit procedures conducted during CSD processing (see section 5.7).

Special codes have been designated to each of these types of responses to facilitate user recognition and data analysis. For instance, "valid skip" codes are set to "6" as the last digit, with any preceding digits set to "9" (for example, code would be "996" for a 3-digit variable). All "don't know" responses end in "7", with any preceding digits set to "9" (for example, "997"). Refusals end in "8", with any preceding digits set to "9" (for example, "998"); and "not stated" values end in 9, with any preceding digits set to "9" also (for example, "999"). Suppressed values are coded to "5" as the last digit, with any preceding digits set to "9" (for example, "995").

5.6 Coding

"Other-specify" items

Data processing also includes the coding of "Other-specify" items, sometimes referred to as "write-in responses". For most questions on the CSD questionnaire, pre-coded answer categories were supplied and the interviewers were trained to assign a respondent's answer to the appropriate category. However, in the event that a respondent's answer could not be easily assigned to an existing category, many questions also allowed the interviewer to enter a long-answer text response in the "Other-specify" category.

All questions with "Other-specify" categories were examined during processing. In the case of questions where 'Other-specify' responses constituted less than 10% of overall responses to the question, coding was not performed and responses were left in 'Other'. A total of 12 questionnaire items were coded for 'other-specify' responses. Based on coding guidelines, many of the long answers provided by respondents for these questions were re-coded back into one of the existing answer categories. Responses that were unique and qualitatively different from existing categories were kept as "Other". Where counts warranted, new categories were created to capture emerging themes in the data that were not reflected in existing categories. Appendix D presents the extra categories added for the 2012 CSD. These will be taken into account when refining the answer categories for future cycles of the survey.

Open-ended questions and standard classifications

A few questions on the 2012 CSD questionnaire were recorded by interviewers in a completely open-ended format. These included questions related to the following:

1. The respondent's main medical condition which causes them the most difficulty or limits their activities the most;
2. Occupation and industry of work; and
3. Major field of post-secondary study, where applicable.

These responses were coded using a combination of automated and interactive (manual) coding procedures. Standardized classification systems were used to code these responses. Appendix D provides details of these classifications.

Coding for all classifications involved a team of experienced coders and quality control supervisors. Subject matter experts in data processing applied additional verification procedures.

5.7 Edits and imputations

After the coding stage of processing, analyses were performed on the data to identify gaps, inconsistencies, extreme outliers and other potential problems in the data. To resolve the problematic data identified, customized edits and imputations were performed.

CSD data were evaluated both for internal consistencies across variables and for external consistencies in relation to the respondents' linked NHS data. With respect to internal consistencies, the data had already benefitted from extensive edits built into the electronic questionnaire.

Examples of electronic edits include several that ensured that responses related to age throughout the questionnaire were not contradictory. For example, edits ensured that the reported age when respondent first experienced limitation in daily activities or when the respondent completed their highest level of schooling was not greater than the age of the respondent at the time of the interview. Another questionnaire edit ensured that the number of hours a respondent reported that they usually worked per week did not exceed 168 (24 hours x 7 days) and prompted interviewers to confirm the hours with respondents whenever reported hours per week exceeded 80.

Post-survey analyses were undertaken to review the frequency with which questionnaire edits were triggered during interviews for all respondents. As none of these counts were significant, further internal consistency editing was not required.

In addition to these analyses, two situations were identified where an error was found on the electronic questionnaire resulting in missing data for some respondents. In both cases, imputation was considered as a potential solution. Imputation involves replacing the missing data with a new value based on values from a donor respondent with similar characteristics or based on values determined through logical determinism.

In the first situation, a subgroup of respondents (n=581) in the module on employment modifications were missed at question EMO_Q02, which asked if certain modifications to their work environment were made available to the respondent. More specifically, due to an error in the programming of the skip pattern, the subgroup of respondents missed were persons with a disability out of the labour force, who have worked since 2007, and who reported that their condition does not completely prevent them from working and would need an accommodation to be able to work. We considered the option of imputing the missing responses for these individuals using donor imputation. However, these persons have a very unique profile, and after closely examining the potential donors, we were unable to find suitable donors with similar characteristics to the respondents. As a result, we kept the responses as missing, rather than imputing responses from donors with very different (employment) profiles.

In the second situation, the responses to questions SNC_Q01D and SNC_Q01E (*"In the past year, did you receive income from the Canada Pension Plan Disability Benefit?"* and *"In the past year, did you receive income from the Quebec Pension Plan Disability Benefit?"*) were imputed for some respondents (n= 466) because their year of birth was missing. These questions were normally asked when a condition on year of birth had been met (YOB $\geq$ 1946). A missing year of birth⁷ resulted in the situation where, for certain respondents, the persons should have answered the questions, but instead the questions were not asked. Since the two possible answers were "yes" and "no," we imputed the responses using two logistic regression models to predict the probability of a "yes" answer to these two questions, respectively. To build and validate these models, we used data from other respondents with a disability who had given a year of birth and who had answered the proper question flow. We then ran the models for the cases with missing data to impute a value, based on the predicted probability of a "yes" answer. Indicators to identify the imputed cases were created for users.

In terms of external edits, analyses were performed comparing CSD and NHS data for those indicators where the concepts across surveys overlapped in meaningful ways and where inconsistencies would have the potential to present analytical problems of interpretability for users. Over 20 indicators were analyzed, primarily in the areas of labour force participation, educational attainment and family status. The basic assumption was that the CSD has more recent data than the NHS and that the NHS variables, in their release version, may have undergone a number of manipulations and imputations. Therefore, the general rule is as follows: when there is an inconsistency, the value of the particular NHS variable is deleted and replaced with the special value "95" to allow users to take it into account during their analyses.

5.8 Derived variables and external NHS-linked variables

In order to facilitate more in-depth analysis of the rich CSD dataset, approximately 100 derived variables (DVs) were created by re-grouping or combining items on the questionnaire. In addition, approximately 200 NHS variables were linked to the final CSD processing file for 2012.

DVs were created to provide users with indicators of disability status and disability type, based on definitions used for the CSD (see section 2.1 for survey definitions of disability). DVs were also created to capture disability severity ratings and classes across disability types. Others included DVs for the use of assistive devices and help needed with daily activities which required combining several items on the questionnaire. Numerous important labour force concepts were captured by DVs, as were sources of income. Finally, several DVs reflected the coding of variables to standard classification systems such as the International Classification of Diseases and the Classification of Instructional Programs.

In constructing derived variables for respondents, if any component question was not answered (that is, had a value of "Don't know", "Refused", "Not stated", or "Valid skip"), the code assigned to the derived variable was labelled "Not stated".

All derived variable names have a "D" in the first character position of the name. For all linked NHS variables, the NHS variable name was preserved as much as possible on the CSD database. Some exceptions applied since CSD variable names are restricted to eight characters whereas NHS variable names sometimes exceeded eight characters.

The 2012 CSD Data Dictionaries identify in detail which variables were derived and the specifications for their derivation (contact Statistics Canada Client Services for details). DVs are listed by theme in Appendix A along with other survey indicators. A complete list of DVs and NHS variables, along with their specifications are provided in the 2012 CSD Data Dictionaries.

⁷ The application used the field corresponding to the date of birth reported directly by the respondent at the time of the interview, but some respondents only reported their age.

6. Weighting and creation of final data files

6.1 Weighting

In a sample survey, each respondent represents not only himself or herself but also other people who have not been sampled. For that reason, each respondent is assigned a weight which indicates the number of people that he or she represents. To maintain data coherence and ensure that the results accurately represent the target population and not just the individuals sampled, that weight must be used to compute all estimates.

There are several steps in calculating the weights for the CSD. The first step is to assign each unit selected for the CSD an initial weight based on the sample design. The initial weight is the inverse of the probability of inclusion. For the CSD, the initial weight is the product of two factors: the NHS weight and the CSD subsampling weight (the inverse of the sampling fraction). Then a number of adjustments are made to the weights to control for non-response and exclusions during collection and to avoid extreme weights in the estimation domains. The final step is to calibrate the survey weights on the NHS estimated totals and make certain adjustments to account for units that were in scope during the May 2011 selection process but out of scope at the time of the survey in 2012. The main steps in the weighting process are described in the subsections below.

Calculation of the initial weights

Initial weights need to be calculated for both the YES sample and NO sample. Since the CSD sample design is based on the NHS design, the initial weight is the product of the NHS final⁸ weight and the inverse of the CSD sampling fraction. The NHS final weight takes into account the NHS sample design, non-response and other corrections. For more details on the NHS weighting strategy, refer to the NHS User Guide, section 4.3.

Hence the CSD initial weight is the product of the adjusted and corrected NHS weight and the inverse of the CSD sampling fraction. The CSD sampling fraction is the size of the sample selected in a stratum divided by the number of units available in the sampling frame in that stratum.

The weight adjustments described in the next subsections only pertain to the YES sample, as there was no collection done for the NO sample, therefore no need to adjust for non contact or non response.

Adjustment for units not sent to collection

The sample selected for the CSD was expanded slightly in anticipation of the exclusion of some units from collection for the following reasons:

- selection of more than three members of the same household;
- selection of persons in households previously selected for other surveys in the North (Canadian Community Health Survey, Labour Force Survey or Survey of Household Spending);
- no telephone number available for a household (outside the NWT);
- no name or date of birth reported in the NHS for a person, so no way to identify the right respondent in the household.

These losses were taken into account in calculating the sample size, and oversampling in some strata was done to compensate for them. Units excluded from collection were thus treated as non-respondents and their weight was redistributed to the other units at the stratum level.

⁸ The NHS weight that was used here is a version that was corrected to avoid excessively large weights.

Non-contact and non-response adjustments

There were two major categories of non-response in the CSD: non-contact, and non-response after contact. These two types of non-response were treated separately, as they constituted two different phenomena. The factors that explain non-contact tended to be related to the characteristics of the household and the mobility of persons, while the factors that explained non-response in a contacted household tended to be related to the characteristics of individuals.

First, units sent to the field were separated into two large groups: units that were contacted, and units that were not contacted. Logistic regression was used to model the probability of being contacted. The explanatory variables for the model came either from the NHS frame or from paradata.⁹ The variables selected for the non-contact model were province, owning or renting the dwelling, number of persons in the household, number of children in the household, number of bedrooms, total household income, 5-year mobility indicator, visible minority group, registered Indian status, class of worker, labour force status, type of occupation, reason for not being able to start work the week before the Census, age group, marital status, consenting to release one's NHS data in 92 years, knowledge of official languages, status in the census family structure, and number of contact attempts.

With this model, for each unit (contacted or not), a probability of being contacted was obtained. Then, response homogeneity classes were formed by grouping units that had similar probabilities of being contacted. An automatic class formation method¹⁰ was used to generate classes that were homogeneous with respect to predicted probability of being contacted and contained a sufficient number of units contacted to avoid excessively large weight adjustment factors. A total of 37 classes were formed, and within each one, the weight of the non-contact units was redistributed to the contacted units.

Next, an adjustment was made for a subset of persons who were contacted but non-respondent, either those who had a disability or health condition that prevents them from responding, or those who completed the DSQ module (and, on the basis of their responses, had a disability) but not the rest of the CSD interview. Since there were very few of these cases (about 700), a relatively simple adjustment was made at the stratum level, redistributing the weight of non-respondents with a disability among respondents with a disability.

The next step was to adjust the weights for other non-respondents (generally refusals). In this case, too, logistic regression was used to model the probability of responding given the fact that contact had been made at the household level. The variables selected for the non-response model were province, age group, presence of children in the household, number of children in the census family, number of maintainers in the household, living in a large urban centre or not, visible minority group, class of worker, occupational group, total household income, consenting to release of one's NHS data in 92 years, knowledge of official languages, number of contact attempts, number of contacts, number of refusals, number of appointments, average number of days between contacts, date of last contact (we divided the collection period into four groups), day of the week of last contact, and time of last contact.

With this non-response model, for each unit (respondent¹¹ or not), a probability of responding given contact was obtained. Response homogeneity classes were formed by grouping units that had similar probabilities of responding. The same procedure was used as for the contact model, which resulted in the formation of 14 classes. Within each class, the weight of the non-respondent units was redistributed among the respondent units.

Out-of-scope units (deaths, institutional admissions, persons who now live outside the country, etc.) were initially considered respondent units, in that we were able to speak with a household member who confirmed the unit's out-of-scope status. Their weight was not set to 0; rather, it was retained since they represented units of the initial population (on May 10, 2011) that were out of scope in the fall of 2012. However, these units are excluded from the analytical file.

⁹ Paradata are collection variables that are available for all the selected units, such as the number of contact attempts, the date, time and duration of each contact, the result of each contact (refusal, appointment, completed interview, etc.).

¹⁰ The SAS procedure PROC FASTCLUS was used here.

¹¹ Note that out-of-scope units were excluded from this adjustment (but included in the non-contact adjustment). Since the household had already been contacted in all these cases, we would have known if the selected person was out of scope. We therefore assumed here that all non-respondents with contact were in-scope, and their weight was redistributed among in-scope respondents only.

Adjustment for extreme weights

Following the non-contact and non-response adjustments, the distribution of respondents' weights was examined to detect the presence of very large weights by province and then by estimation domain. Some adjustment factors may have generated very large weights for some individuals compared with others in some domains, which could have a detrimental effect on the estimates and their variance. The sigma-gap method was used to detect these extreme weights first within each province and then within each estimation domain. An example of how the sigma-gap method can be applied is given in Bernier and Nobrega (1998).¹² As used here, the sigma-gap method is intended to detect large gaps between successive weights sorted in ascending order (when they are greater than the median). When an excessively large gap is found between two successive weights, the larger of the two weights and all subsequent weights are classified as outliers. To assess the size of a gap between two weights, it was compared with a certain number of standard deviations of the distribution of all weights. For the CSD, gaps between weights that were three times the distribution's standard deviation within each province were identified. The choice to use three standard deviations was made because that corresponded to the gap that would have been used to identify outlier weights with a manual process. All the weights identified as outliers were set to the province's highest non-outlier value. The same procedure was then followed for the estimation domains. The resulting weight reduction was offset in the post-stratification step, as described below.

Estimation domain jumpers

Before proceeding to post-stratification, we had to account for cases in which the age derived from the date of birth reported in the CSD did not match the age group to which the case was assigned on the basis of the sampling frame. This situation is possible since the date of birth in the NHS frame may have been missing and been imputed, or it may have been incorrectly reported or captured. We found just under 300 cases where this had occurred. We compared their weights following the adjustment for extreme weights in the previous step with the distribution of weights in their new domain. When the individual's weight fell within the range of weights in the new domain, it was retained with no change. On the other hand, if it fell outside the range of weights in the new domain, it was changed to the new domain's minimum value (if it was below the range) or maximum value (if it was above the range). In this step, we adjusted the weight of 35 individuals in the CSD sample.

¹² Bernier, J., and Nobrega, K. (1998). "Outlier detection in asymmetric samples: A comparison of an inter-quartile range method and a variation of a sigma-gap method". Annual meeting of the Statistical Society of Canada, June 1998.

Post-stratification and out-of-scope adjustment

Two separate post-stratification steps were needed to obtain the final weights. These steps were done for both the YES sample and NO sample.

First post-stratification

The first post-stratification involved adjusting the weights of CSD respondents (out-of-scope cases, respondents with a disability, and respondents without a disability) to obtain the same totals as in the NHS for the YES population by province, age group, sex and severity. Severity refers to the severity reported in the NHS: either "often" or "sometimes". Post-stratification was performed on the following age groups: 15-19, 20-24, 25-29, ..., 60-64, 65-79, 75+.

The same post-stratification was carried out on the NO sample starting from the initial weights computed earlier.

The first post-stratification was performed independently for the YES and NO populations. The control totals were calculated from the NHS dissemination database for the population aged 15 and over living in private dwellings, excluding reserves.

Second post-stratification

The main purpose of the second post-stratification was to adjust the weights of the NO sample in order to estimate the proportion of the population that was out of scope in the fall of 2012. Nonetheless, this step was performed on both the YES sample and NO sample. Since CSD collection took place about a year and a half after the NHS, a large number of out-of-scope units were found in the YES sample, representing about 235,000 individuals in the YES population. Because these out-of-scope cases must be excluded from the disability rates, it is important to try to remove the out-of-scope cases from the denominator (which consists of the YES and NO populations) as well. Otherwise, we may underestimate the disability rates. As the NO sample was not sent to collection, we had to find an indirect method of estimating and excluding the out-of-scope cases. The strategy used to estimate the number of out-of-scope cases in the NO population, as we will see, depends on the type of out-of-scope case and the available information.

First, we calibrated the weights of the YES and NO samples on the population estimates¹³ produced by the Demography Division, adjusted for net undercoverage on May 10, 2011, excluding the under-15 age group, Indian reserves, collective dwellings, members of the Canadian Forces, and visitors to Canada. This calibration was necessary because the death and emigration totals we will use later to estimate the number of out-of-scope cases in the NO population relate to the total population, not just the enumerated population. More than 65% of the out-of-scope cases observed in the CSD were deaths. The adjustment for this type of out-of-scope case had to be as precise as possible.

Second, we made the various adjustments required to exclude units that would have been out of scope in the fall of 2012 from the NO population. The aim here is not to identify the individuals in the NO sample who are out of scope but to reduce the sum of the NO sample's weights so that they match the population totals excluding the out-of-scope cases.

These adjustments were carried out separately for different types of out-of-scope cases, and using different data sources, as described below.

Estimates for the various types of out-of-scope cases in the YES population are presented in the table below.

¹³ When weighting was done for the CSD, the Demography Division was still producing its population estimates with a model based on the 2006 Census. The data needed to adjust the 2011 Census data for net undercoverage were not yet available. We nevertheless decided that it was better to make this adjustment to avoid underestimating the CSD disability rates.

Table 6.1
Estimated number of out-of-scope cases among the YES population

Type of out-of-scope	Unweighted	Weighted	Weighted
	number		%
Deaths	939	153,570	65.40
Emigrants	66	15,870	6.80
Institutional admissions	349	57,450	24.50
Moved to an Indian reserve	4	120	0.05
Other	62	7,800	3.30
TOTAL	1,420	234,810	100.00

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Deaths and emigrants (between May 10, 2011, and CSD collection)

The Demography Division produced cumulative totals of deaths and emigrants between May 10, 2011, and November 15, 2012 (the mid-point of CSD collection), by province/territory, age group and sex. These totals were adjusted to exclude people under age 15, Indian reserves, collective dwellings, members of the Canadian Forces and visitors to Canada. Since these totals cover the entire population (the YES and NO populations), we were able to subtract the number of deaths estimated for the YES population with the CSD sample (using the weights obtained following post-stratification on the Demography Division totals, as described above) and thus derive the number of deaths in the NO population. Then we were able to adjust the sampling weight of the NO population downward to produce an estimate of the population that was still alive and still in Canada in the fall of 2012.

Table 6.2
Estimated number of deaths and emigrants between 2011 Census day and 2012 CSD collection

	CSD estimate (YES population)	Demography Division estimate (total population)	Indirect estimate (NO population)
Deaths	153,570	338,520	184,950
Emigrants	15,870	70,290	54,420

Sources: Statistics Canada, Canadian Survey on Disability, 2012 and Population Estimates Program.

The downward adjustment of the weights of the NO population was carried out for deaths at the level of the provinces cross-classified by two age groups (15-64 and 65+). For emigration, the adjustment was made by age group only (15-24, 25-44 and 45+). It was important for the adjustment to be made as precisely as possible, but without creating inconsistencies in the adjustments to be carried out.

Persons living in an institution after May 10, 2011

This type of out-of-scope case is more difficult to estimate for the NO population. There is no accurate data source that we can use to estimate the number of people admitted to institutions since the last census. In addition, we know that the proportion of people admitted to institutions in the YES population should be larger than the proportion in the NO population, but we do not know how much larger.

We obtained estimates of institutional admissions from the Survey of Labour and Income Dynamics (SLID, Panel 7) for the total population aged 15 and over (with the same exclusions of reserves, collective dwellings, etc., as the CSD, but also excluding the territories). These estimates were provided to us by age group (15-64, 65-74, 75+), and they cover a period of a year and a half, which is comparable to the gap between the Census and the CSD. Since the SLID has no estimates for the territories, it was impossible to estimate the number of institutional admissions in the territories' NO population. An indirect estimate of the institutional admissions in the NO population was obtained by subtracting the number of institutional admissions estimated for the YES population in the CSD from the SLID's estimated total for the entire population (as was done for deaths and emigration).

The number of institutional admissions in the NO population was negligible for the first two age groups (15-24 and 65-74). We therefore performed an adjustment only for the last age group in the NO population. We decreased the weight of the NO population aged 75 and over to reduce the population by just under 26,000 in the 10 provinces.

Persons who moved to an Indian reserve after May 10, 2011

The number of cases in the YES population that were out of scope because they moved to a reserve was so small that we saw no need to make an adjustment in the NO population.

Other out-of-scope cases

For the other types of out-of-scope cases—members of the Canadian Forces, visitors to Canada (incorrect classification in the NHS) and invalid age—we had to assume that the proportion for the NO population would be the same as the proportion estimated for the YES population. This assumption is less realistic for members of the Canadian Forces, but this adjustment for out-of-scope cases is very small.

The total number of members of the YES population who fall under these three types of out-of-scope cases makes up about 0.13% of the YES population.¹⁴ We therefore assumed that 0.13% of the NO population also fell under these types of out-of-scope cases, and we reduced the weights of the NO population in proportion to the observed figures for the YES population in the following three age groups: 15-24, 25-44, 45+.

It is important to keep in mind that the out-of-scope cases in the NO population were excluded by adjusting the weight of the NO population downward to compensate for the losses. For the YES population, on the other hand, we simply excluded from the analytical file those individuals who were identified as out-of-scope cases and whose sampling weight enabled us to estimate the number of persons they initially represented in the YES population.

Weighted population counts

Table 6.3 shows counts for the YES population covered by the survey, that is, persons who reported an activity limitation in the NHS filter questions, after adjustment for net undercoverage and exclusion of units that were out of scope in the fall of 2012.

Table 6.3
YES population from the NHS covered by the CSD, and number of persons with a disability, by province and territory

Province/Territory	YES Population covered	Persons with a disability
	number	
Newfoundland and Labrador	85,130	59,300
Prince Edward Island	26,370	18,840
Nova Scotia	199,490	143,760
New Brunswick	145,300	99,450
Quebec	956,330	616,740
Ontario	2,399,620	1,651,620
Manitoba	214,640	145,270
Saskatchewan	172,330	116,640
Alberta	573,660	369,190
British Columbia	891,330	546,760
Yukon	6,300	4,070
Northwest Territories	3,900	2,740
Nunavut	2,900	1,540
Canada	5,677,290	3,775,910

Source: Statistics Canada, Canadian Survey on Disability, 2012.

¹⁴ The proportion 0.13% is obtained by dividing 7,739 by the total of the CSD weights after the second post-stratification.

The next table shows weighted counts of the population covered by the NO sample, following adjustments for net undercoverage and exclusion of units that were out of scope in the fall of 2012.

Table 6.4
NO population from the NHS covered by the NO sample, by province and territory

Province/Territory	Population covered
	number
Newfoundland and Labrador	335,850
Prince Edward Island	91,070
Nova Scotia	565,610
New Brunswick	461,520
Quebec	5,480,590
Ontario	8,328,280
Manitoba	715,010
Saskatchewan	606,820
Alberta	2,371,490
British Columbia	2,811,680
Yukon	22,060
Northwest Territories	29,470
Nunavut	19,450
Canada	21,838,910

Source: Statistics Canada, Canadian Survey on Disability, 2012.

6.2 File structure and content

We created two analytical files for the CSD: a file with persons who have a disability, and a file with persons who do not have a disability. Depending on the type of analysis required, users will have to use either the disabled persons file only or both files together.

The **disabled persons** file contains persons selected for the CSD who, according to the definition of disability used in the CSD, are considered to have a disability. This file is the more comprehensive of the two. It contains all CSD data and several variables from the NHS. Any analysis that deals exclusively with disabled persons can be done with this file alone.

The non-disabled persons file contains three groups of people: two from the CSD's YES sample, and the third from the NO sample. The three groups are as follows:

- Persons interviewed for the CSD who reported being limited only "rarely" with "no difficulty" or "some difficulty". For the purposes of the CSD interview, these persons were deemed to potentially have a disability and therefore had to answer all the questions in the CSD. However, they were excluded from the final definition of disability.
- Persons interviewed for the CSD who reported that they were "never" limited (false positives). For the purposes of the CSD interview, these persons were deemed not to have a disability and therefore did not have to answer the rest of the questions in the CSD.
- Persons from the NO sample. They reported no activity limitations in the NHS, were not sent to collection, and were automatically deemed not to have a disability.

Hence, the non-disabled persons file has different content depending on the group of people involved. For the persons in group (a), we have the same content as for the persons with a disability: all the CSD data and several variables from the NHS. For the persons in group (b), we have only the data from the CSD's DSQ module, since the interview was terminated immediately after that module. However, we also have the NHS variables. For the persons in group (c), we have only the NHS variables, since no CSD collection was done for those units.

The non-disabled persons file should be used together with the disabled persons file for two types of analysis: calculation of disability rates, since the denominator must include both persons with a disability and persons without a disability, and comparison of the NHS characteristics¹⁵ of persons with a disability and persons without a disability.

To distinguish between the various groups in the files, we created a derived variable, CSDPOPFL, which takes a value of 1 for persons with a disability, 2 for group (a) persons without a disability, 3 for group (b) persons without a disability, and 4 for group (c) persons without a disability.

To make the two files easier to use, they both have the same variables. However, some variables will have missing values in the non-disabled persons file, because we do not have all the information for some groups. The table below summarizes the contents of the various files and the various groups, for the disabled persons file and the non-disabled persons file.

Table 6.5
Available content for various groups of persons in the CSD analytical files

	CSDPOPFL	Demographic variables	DSQ	Other CSD content and derived variables	NHS variables	Final weight and bootstrap ¹ weights
Persons WITH a disability	1	✓	✓	✓	✓	✓
Persons WITHOUT a disability (group (a))	2	✓	✓	✓	✓	✓
Persons WITHOUT a disability (group (b))	3	✓	✓		✓	✓
Persons WITHOUT a disability (group (c))	4	✓			✓	✓

✓ content is available

1. For a description of bootstrap weights, see section 7.2.

Source: Statistics Canada, Canadian Survey on Disability, 2012.

A note on reference periods

When calculating disability rates or comparing the characteristics of persons with a disability and persons without a disability, the **reference date is May 10, 2011**. However, when one is only interested in persons with a disability, one will work with the CSD data collected and measured in the fall of 2012. In this case, the **reference period will be the Fall of 2012**. This is equivalent to viewing the CSD as a sort of longitudinal survey where the first wave of data was collected in 2011 through the NHS (the initial population), for both the YES population and the NO population, and where the second wave was collected in 2012 through the CSD, for the subset of persons with a disability.

In other words, the CSD's persons with a disability are individuals who reported an activity limitation in the NHS on May 10, 2011, and a disability in the CSD in 2012. Hence, the CSD's characteristics for persons with a disability are based on **2012 information about a population defined in 2011**.

¹⁵ NHS characteristics, because for most people without a disability, we have only those variables (the NO sample).

6.3 Final datasets and data dictionaries

Final data files included the following:

- Final processing file
- Analytical files for use in Research Data Centres

The final processing file is an in-house file that includes a number of temporary variables used exclusively for processing purposes. The Analytical files are dissemination files processed further for release purposes. The analytical files are distributed in Research Data Centres across Canada. They are also used at Statistics Canada to produce data tables in response to client requests. Dissemination files are scheduled for distribution following the CSD release day of December 3, 2013 (see Chapter 9 for dissemination details).

In order to transform the final, cleaned processing file to final Analytical data files for researchers, a series of actions were performed. First, steps were taken for the enhanced protection of respondent confidentiality. Next, person-weights were added to the files. Weighting is described in more detail in Section 6.1. Finally, all temporary variables or variables used exclusively for processing purposes were removed from the files.

Accompanying the 2012 CSD Analytical files are the following supporting documents:

- the record layout,
- SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences) and Stata syntax to load the files, and
- metadata in the form of a data dictionary for each Analytical file that describe every variable and provide weighted and unweighted frequency counts.

6.4 Guidelines for analysis

A User Guide has been created for the Analytical files. This guide provides detailed step-by-step instructions for using the CSD data files. The User Guide includes guidelines for tabulation and statistical analysis, how to apply the necessary weights to the data, information on software packages available and guidelines for the release of data, such as rounding rules. The process of estimating the reliability of estimates, both quantitative and qualitative, is covered in detail. In addition, detailed data dictionaries are available covering all variables available.

7. Data quality

7.1 Overview of data quality evaluation

The objective of the Canadian Survey on Disability is to produce quality estimates on the type and severity of disabilities of Canadians aged 15 years and over (as of May 10, 2011) as well as on a variety of other important indicators of the experiences and challenges of persons with disabilities. This chapter reviews the quality of the data for this survey.

Sections 7.2 and 7.3 below explain the two types of errors that occur in surveys - sampling and non-sampling errors. Each type of error is evaluated in the context of the CSD. Sampling error is the difference between the data obtained from the survey sample and the data that would have resulted from a complete census of the entire population taken under similar conditions. Thus, sampling error can be described as differences arising from sample-to-sample variability. Non-sampling errors are all other errors that are unrelated to sampling. Non-sampling errors can occur at any stage of the survey process, and include non-response for the survey as well as errors introduced during data collection or computer processing.

This chapter describes the various measures adopted to prevent errors from occurring wherever possible and to adjust for any errors found throughout the different stages of the CSD. Areas of caution for interpreting CSD data are noted. Readers may also refer to the National Household Survey User Guide for related information on data quality.

7.2 Sampling errors and bootstrap method

The estimates that can be produced with this survey are based on a sample of individuals. Somewhat different estimates might have been obtained if we had conducted a complete census with the same questionnaires, interviewers, supervisors, processing methods and so on, as those actually used. The difference between an estimate derived from the sample and an estimate based on a comprehensive enumeration under similar conditions is known as the estimate's "sampling error".

To produce estimates of the sampling error for statistics produced from the CSD, we used a particular type of bootstrap method. Several bootstrap methods exist in the literature, but none was appropriate for the CSD's complex sample design. The following characteristics of the sample design make it difficult to estimate the sampling errors:

- A three-phase design in which households (or dwellings) are selected in the first two phases, and individuals in the third phase. In the first phase, a random sample of 4.5 million households stratified by collection unit (CU) was selected for the NHS. In mid-July 2011, a second-phase subsample of 400,000 households from the 1.2 million that had not yet responded was selected as part of the non-response follow-up (NRFU) operation. In the third phase, a sample of some 45,500 individuals with activity limitations according to the NHS was selected for the CSD.
- The sampling fraction of the first-phase sample (NHS) is non-negligible (about 1/3 in the N1 regions), and the sampling fraction of the CSD is rather high in some strata.
- The CSD strata (combinations of estimation domains, N1 or N2 regions, initial respondent vs. NRFU respondent, often limited vs. sometimes limited) are non-nested within the NHS strata (CUs or groups of CUs).
- The method used has to be flexible enough to produce standard statistics such as proportions, totals, averages and ratios, as well as more sophisticated statistics, including percentiles and logistic regression coefficients.
- For the purposes of calculating the sampling error, NRFU respondents were treated as a third-phase sample, in which a dwelling's probability of inclusion was equal to its probability of responding, independently for each dwelling.

As previously mentioned, several bootstrap methods exist in the literature for one-phase sampling. The most common method, known as "with replacement" bootstrap, involves selecting M with-replacement subsamples from the main sample and producing estimates for each subsample. The bootstrap variance estimate is then calculated as a function of the squared differences between the estimates from each of the M bootstrap subsamples and the estimate from the survey sample.

The use of bootstrap weights makes calculating the variance much simpler. For each subsample (bootstrap replicate), the initial sampling weights are adjusted for bootstrap subsampling, which produces what is known as "initial bootstrap weights". Since each bootstrap sample is the result of selecting units with replacement, a unit can be selected in a particular bootstrap sample more than once. It can be shown that the bootstrap weights are a function of the initial weight of the observation multiplied by what is referred to as the unit's "multiplicity" in the bootstrap sample, which is the number of times the unit is selected in the bootstrap sample. A unit's multiplicity in the bootstrap sample is a random variable that has what is known as a "multinomial distribution". Hence, the bootstrap weights can be seen as the product of the initial sampling weights and a random adjustment factor (in this case, a function of the unit's multiplicity). Once initial bootstrap weights have been calculated, all weight adjustments applied to the initial sampling weights were applied to the initial bootstrap weights to obtain the final bootstrap weights, which capture the variance associated not only with the particular sample design but also with all weight adjustments applied to the full sample to derive the final weights.

In 2006, the Aboriginal Peoples Survey (APS) developed a general bootstrap method for two-phase sampling (Langlet, Beaumont and Lavallée, 2008). That method could not be used for the Participation and Activity Limitation Survey (PALS) because it was still being developed when the data were released. For 2011, the two-phase method used in the 2006 APS was adapted for the new complexities of the sample design. A review of the 2006 method follows.

As mentioned previously, the bootstrap weights can be seen as the product of the initial sampling weights and a random adjustment factor. This is the underlying idea of the general bootstrap method. In the case of a two-phase sample, the variance can be decomposed into two components, each one associated with a sampling phase. The two-phase general bootstrap method generates a random adjustment factor for each phase of sampling. In this case, the initial bootstrap weight of a given unit is the product of its initial sampling weight and the two random adjustment factors.

In 2011, for variance estimation, the NHS was considered to have two additional phases: the NRFU subsample and the NRFU subsample non-response. To take these additional phases into account, the three phases of the NHS were combined into a single phase so that the two-phase general bootstrap method (an NHS phase and a CSD phase) could be used. In the general bootstrap method for two-phase designs, it can be shown that the random adjustment factors depend on the single and double inclusion probabilities associated with each phase. To combine the three phases into a single phase, the single and double inclusion probabilities for the three phases of the NHS need to be combined. The single and double inclusion probabilities of the three phases combined are given by the product of the single and double inclusion probabilities of each of the three phases. The details of the methodology used are provided in Haddou (2013).¹⁶

Once the three phases of the NHS have been combined into a single phase, the general bootstrap method for two-phase sampling was used, which involved the derivation of two sets of random adjustment factors, one for each phase.

There is a major advantage in having two sets of random adjustment factors. The first set of adjustment factors can be used for estimates based on the first phase only, i.e., estimates based on the NHS sample. These estimates are used when the weights are adjusted to the NHS totals during post-stratification (Section 6.1). This produces variable NHS totals for each bootstrap sample, which reflects the fact that the NHS totals used are based on a sample and not on known fixed totals.

For the CSD, 1,000 sets of bootstrap weights were generated using the method described above. The method used is slightly biased in that it slightly overestimates the variance. The extent of the overestimation is considered negligible for the CSD. The method can also produce negative bootstrap weights. To overcome this problem, a transformation was performed on the bootstrap weights to reduce their variability. Consequently, the variance calculated with these transformed bootstrap weights has to be multiplied by a factor which is a function of a certain parameter, known as ϕ . The parameter's value is chosen as the smallest integer

16 Haddou, M. (2013). "Bootstrap Variance Estimation Specifications - Aboriginal Peoples Survey". Internal document, January 2013.

that makes all bootstrap weights positive. For the CSD, this factor is 4. The variances calculated from the transformed bootstrap weights must therefore be multiplied by $4^2 = 16$. Similarly, the coefficients of variation obtained (square root of the variance divided by the estimate itself) have to be multiplied by 4. However, most software applications that produce sampling error estimates from bootstrap weights have an option to specify this adjustment factor, so that the correct variance estimate is obtained without the extra step of multiplying by the constant.

It is extremely important to use the appropriate multiplicative factor for any estimate of sampling error such as variance, standard error or CV. Omission of this multiplicative factor will lead to erroneous results and conclusions. This factor is often specified as the "Fay adjustment factor" in software applications that produce sampling error estimates from bootstrap weights.

For examples of procedures using the Fay adjustment factor, see the 2012 Canadian Survey on Disability Data User Guide.

The measure of sampling error used for the CSD is the coefficient of variation (CV) of the estimate, which is the standard error of the estimate divided by the estimate itself. For this survey, when the CV of an estimate is greater than 16.5% but less than or equal to 33.3%, the estimate is accompanied by the letter "E" to indicate that the data should be used with caution. When the CV of an estimate is greater than 33.3%, or if an estimate is based on 10 units or less, the cell estimate is replaced by the letter "F" to indicate that the entry is suppressed for reliability reasons.

7.3 Non-sampling errors

Besides sampling, a number of factors at almost every phase of a survey can cause errors in survey results. Respondents may misunderstand the questions and answer them inaccurately, responses may be entered incorrectly during data capture and errors may be introduced in the processing and tabulation of data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates drawn from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Thus, much time and effort was devoted to reduce non-sampling errors in the survey. At the stage of content development, extensive activities were undertaken to develop questions that would be well-understood by respondents. The new questionnaire was tested thoroughly during several rounds of qualitative testing. In addition, many initiatives were taken in the field to encourage participation and reduce the number of non-response cases. Also important were the numerous quality assurance measures applied at the stages of data collection, coding and processing to verify and correct errors in the data. Weighting adjustments were used where appropriate to correct potential bias.

Non-sampling errors arise primarily from the following sources: coverage, non-response, measurement and processing. For each of these areas, the following sections discuss the various measures used to minimize and correct error.

Coverage errors

Coverage errors occur when the sampled population excludes people intended to be in the target population. Because the CSD is an extension of the 2011 NHS, it inherits the coverage problems of that survey, which in turn inherits the coverage problems of the 2011 Census. For more information about coverage errors in the census please see the *Final estimates of 2011 Census coverage* on Statistics Canada's website. For more information about the quality of NHS data, see the *NHS User Guide* on Statistics Canada's website.

Non-response errors

Non-response errors result from a failure to collect complete information on all units in the selected sample. Non-response produces errors in the survey estimates in two ways. First, non-respondents often have different characteristics from respondents, which can result in biased survey estimates if non-response is

not corrected properly. In this case, the larger the non-response rate, the larger the bias may be. Secondly, if non response is higher than expected, it reduces the effective size of the sample. As a result, the precision of the estimates decreases (the sampling error on the estimates will increase). This second aspect can be overcome by selecting a larger sample size initially. However, this will not reduce the potential bias in the estimates.

The scope of non-response varies. One level of non-response is item non-response, where the respondent does not respond to one or more questions, but has completed a significant pre-defined portion of the overall questionnaire. Generally, the extent of partial non-response was small in the CSD as a result of extensive qualitative reviews and testing of questionnaire items. There is also total non-response when the person selected to participate in the survey could not be contacted or did not participate once contacted. Weights of respondents were inflated in order to compensate for those who did not respond as described in Section 6.1.

To reduce the number of non-response cases, many initiatives were undertaken prior to and during data collection.

The Statistics Canada website contained a CSD web page which included a series of questions and answers for respondents, as well as general information about the survey. In the months leading up to the survey, a special link to Statistics Canada's website was made available on the websites of several organizations for persons with a disability, so that they could access information about the upcoming CSD. Prior to collection, a brochure was distributed to each selected respondent along with an introductory letter which provided an overview of the survey and explained the importance of participating. A small leaflet was also provided in Braille.

In addition, in-depth interviewer training was conducted. The interviewers were trained by experienced Statistics Canada training staff. In conjunction with the training, detailed interviewer manuals were provided as a reference. Furthermore, all of the interviewers were under the direction of interviewer supervisors, who oversaw activities in the field. Rigorous efforts to reach non-respondents through call-backs and follow-ups were also made by senior interviewers to encourage respondents to participate in the survey.

A table of final response rates obtained for the 2012 CSD is provided in Section 4.2 of this guide. The overall response rate for the survey was 74.6%. Response rates were highest in the older age groups.

Measurement errors

Measurement errors occur when the response provided differs from the real value. Such errors may be attributable to the respondent, the interviewer, the questionnaire, the collection method or the respondent's record-keeping system. Extensive efforts were made for the 2012 CSD to develop questions which would be understood, relevant and sensitive for respondents.

Several rounds of qualitative testing were done for the CSD, and in particular for the new Disability Screening Questions (DSQ). Qualitative testing was carried out by Statistics Canada's Questionnaire Design Resource Centre (QDRC). To minimize measurement error, adjustments were made to question wording and flows based on those results.

Many other measures were also taken to specifically reduce measurement error, including the use of skilled interviewers, extensive training of interviewers with respect to the survey procedures and content, and observation and monitoring of interviewers to detect problems of questionnaire design or misunderstanding of instructions.

Processing errors

Processing errors may occur at various stages including programming of the CATI application, data capture by the interviewer, coding and editing. Quality control procedures were applied to every stage of the data processing to minimize this type of error. Interviews for the CSD were done by Computed Assisted Interviewing and several edits were implemented in the system so that unusual values could be confirmed by respondents during the interview¹⁷ and errors corrected on the spot.

¹⁷ The only exception to this is for the personal interviews done in the Northwest Territories using paper questionnaires that were later on captured in the CATI application at the regional office.

At the data processing stage, a detailed set of procedures and edit rules were used to identify and correct any inconsistencies between the responses provided. For every step of data cleaning, a set of thorough, systematized procedures were developed to assess the quality of every variable on file and to make corrections to every error found. A snapshot of the output files was taken at each step and verification was made comparing files at the current and previous step. The programming of all edit rules were exhaustively tested before being applied to the data. Examples of data processing verification included the review of all question flows, including very complex sequences, to ensure skip values were accurately assigned and distinguished from different types of missing values; quality control double-coding of 'other-specify' responses; experienced supervision of coding to standardized classifications; and the review of all derived variables against their component variables to ensure accurate programming of derivation logic, including very complex derivations. See processing chapter of this guide for more details.

8. Differences between the 2012 Canadian Survey on Disability (CSD) and the 2006 Participation and Activity Limitation Survey (PALS)

The 2012 Canadian Survey on Disability (CSD) includes a set of disability screening questions that were used for the first time to identify persons with a disability in Canada. Although some data users may be seeking to compare the prevalence of disability between surveys, and particularly with the CSD's predecessor – the 2006 Participation and Activity Limitation Survey (PALS) – there are many reasons why this is not possible.

The concepts and methods used to measure disability in the 2012 CSD represent a significant change from those used in the 2006 PALS. The most important change is that the two surveys used a different definition of disability. In the CSD, the definition was applied by using the new set of disability screening questions (DSQ). These screening questions reflect a fuller implementation of the social model of disability, greater consistency in disability identification by type, and improved coverage of the full range of disability types, especially mental/psychological and cognitive (learning and memory) disabilities¹⁸. Differences are discussed in more detail below.

Because of the major differences in concepts and methods between the 2006 PALS and the 2012 CSD, it is neither possible nor recommended to compare the prevalence of disability over time between these two sources.

8.1 New method of screening for disability

In contrast to the PALS screening questions which used a hybrid approach—a social model for identifying some types of disabilities and a medical model for other types—the CSD screening questions (DSQ) were designed to provide greater consistency in disability identification by type.

Based on their responses to the DSQ, respondents are identified as having a disability only if their daily activities are limited¹⁹ as a result of an impairment or difficulty with particular tasks.

The CSD (by adopting the DSQ) allows respondents to determine whether they face activity limitations as a result of these difficulties or impairments. Some people who indicate that they have some difficulty with certain tasks or have an impairment of some type go on to indicate that this never interferes with their daily activities. In PALS, these individuals were considered to have a disability, but in the CSD, they are not.

This change will have the greatest impact on the identification of persons with sensory and physical disabilities because the PALS identified disabilities in these areas solely on the basis of an indication of some difficulty. At the same time, for certain non-physical disability types, PALS is closer to the CSD because it did have the added requirement of a limitation of activities.

Other changes to screening questions may also have an impact on results. For example, the questions regarding mental/psychological disabilities have been altered somewhat by including examples of the more prevalent conditions (such as depression, anxiety, and bipolar disorder) and excluding examples of less prevalent conditions which are also more highly stigmatized (such as schizophrenia). Changes were also made to the list of examples in questions pertaining to learning disabilities and memory disabilities.

An additional difference between the surveys involves the identification of communication disabilities which was done in PALS but not in the CSD. For the DSQ, no question could be found in successive rounds of qualitative testing to properly identify persons with communication disabilities. Most iterations of a question to identify this small group (including the PALS question itself) yielded difficulties due to people having neither English nor French as their first language or to cultural difficulties (for example, not understanding colloquial references). As well, the advent of social media as a form of communication appears to have added new complexities to the concept of communication for the Canadian population.

¹⁸ This is true of the DSQ module on its own; however some constraints are associated with it being administered as a component of the 2012 CSD, a post-censal survey. See Section 8.3 for more detail.

¹⁹ The only exception to this is for developmental disabilities where a person is considered to be disabled if the respondent has been diagnosed with this condition.

Finally, changes were made to the concept of "agility" used by PALS. In the DSQ, this type of disability was split into two types: flexibility and dexterity, since qualitative tests showed that people find that these two tasks are quite different from each other and relate to different underlying conditions. This split was considered an improvement in the identification of different physical disabilities and was also in response to requests made by disability data users and by the Employment and Social Development Canada Persons with Disabilities Technical Advisory Group (TAG).

8.2 Other changes to CSD content

Just as the prevalence of disability and the prevalence of certain types of disabilities cannot be compared across surveys, data regarding other content of the CSD cannot be compared with PALS either. The content of the CSD has been streamlined and updated to a large extent. Some of the content from the PALS was cut due to operational constraints; however, every effort was made to ensure that most of the cuts were restricted to content that had been less informative and less utilized.

Survey questions were also updated to better reflect current realities and to correct known weaknesses in the PALS. For example, the section on aids and assistive devices has undergone major changes. Many of the items contained within the PALS were considered out of date in terms of their current usage by people with disabilities. Similarly, new items were added to better reflect technological advancements that have happened since the PALS questions on aids and assistive devices were originally developed in the late 1990s.

Efforts were also made to streamline the method by which CSD respondents were asked about their requirements and unmet needs. For example, questions regarding the need for and use of household fixtures (such as grab bars, etc.) by persons with certain types of physical disabilities are now combined with other aids and devices for the same disability type. The PALS separated questions about aids and devices that were portable and attached to the persons themselves from those attached to homes. This separation was unnatural for many individuals who wanted to indicate a need for items such as grab bars in the earlier section on aids/devices (often reported under "other") and then reported them again in a later section when specifically prompted, possibly leading to some double counting.

These changes to questionnaire wording and flows mean that comparisons should not be made between the PALS and the CSD data.

8.3 NHS filter questions

While the CSD and the DSQ are considered to be a big step forward in improving the measurement of disability using the social model, it should be noted that the CSD sample was pre-filtered using the same filter questions on the 2011 National Household Survey (NHS) as those used on the 2006 Census long form for the PALS. Follow up studies have shown that these filter questions do not adequately identify people with mental/psychological or cognitive disabilities. This means that the CSD continues to have some of the weaknesses that the PALS had with respect to undercoverage of some disability types. Nevertheless, of those screened in by the NHS, the new method of screening on the CSD will help improve the identification of persons with mental/psychological, cognitive and "other" types of disabilities because they can now be better identified.

8.4 Change in lag time from the filtering survey

As mentioned above, both PALS and CSD derived their sample frames from the answers to the long-form Census in 2006 and the NHS in 2011. The PALS questionnaire was administered between six and nine months after Census data were collected. The CSD, on the other hand, was in the field 16 to 20 months after the NHS data were collected. This difference in lag time not only made it more difficult to track selected respondents who had moved, but it also increased the possibility that a respondent who had reported an activity limitation at the time of the NHS may no longer have a disability, may have been institutionalized or may have died during that time (see Section 8.5). In addition, some information linked from the NHS to the CSD file (for example, information on income) may have changed during the period of time between the two surveys.

8.5 Other methodological changes

The CSD sampling frame was built based on answers to the 2011 NHS, while that of PALS was taken from the 2006 long-form Census. Although every effort was made by the NHS to minimize impacts due to a lower response rate, the CSD results may have been impacted by this change. For more details on the non-response follow-up and weighting adjustments for the NHS, please refer to Chapter 6.

A second methodological difference of the CSD compared to the PALS involved a change in the weighting strategy applied to the CSD to compensate for the longer time lag between the NHS collection and the CSD collection. As mentioned earlier, this time lag increased the likelihood of non-response in the CSD due to death or institutionalisation. As many of these cases may have been persons with a disability, it was important to ensure that disability prevalence not be underestimated. The weights of the population who said NO to the NHS filter questions were, therefore, adjusted to take into account deaths and institutionalizations that would have occurred between the NHS collection and CSD collection. This required a calibration of the weights to population estimates adjusted for net undercoverage, which was not done in the 2006 PALS. For more details on the weighting used in the CSD, please refer to Chapter 6.

8.6 Summary and recommendation

As discussed above, the main differences between the PALS and the CSD can be summarized as follows:

- The definition of disability used in the CSD is different from PALS. The CSD has adopted the newly developed DSQ which is being used for the first time to identify disability in Canada.
- Screening questions in the CSD more closely reflect a social model of disability than do the PALS screening questions. They are also consistent across all types of disabilities, unlike the PALS questions.
- Questionnaire content has been streamlined and updated to reflect current technology and to correct weaknesses in question wording.
- The longer lag time between the NHS and CSD follow-up increased the possibility that selected respondents no longer had a disability, were institutionalized or died during that time. This required a different method for calibration of weights which was not done in the 2006 PALS.
- Finally, the sampling frame for the CSD was derived from the 2011 NHS rather than the 2006 Census and so CSD results may have been impacted by this change.

All of these changes should be assumed to affect comparability of the surveys. Comparison of CSD data to PALS data is, therefore, neither possible nor recommended.

9. Data dissemination

9.1 Data products and services

Data for the 2012 Canadian Survey on Disability (CSD) were released publicly on December 3, 2013. The CSD release included a brief Fact Sheet about disability in Canada and a set of data tables on disability rates for adults in Canada, by age and sex, for each of the provinces and territories. Tables also included data on the types and severity of disabilities. These items are available to the public free of charge on Statistics Canada's website.

Starting in 2014, researchers across the country will be able to conduct in-depth analyses using the CSD Analytical data files housed at Statistics Canada's Research Data Centres (RDCs). In order to access the files, researchers must undergo a research and ethics committee review for approval. Their use of the data must be conducted according to Statistics Canada policies, guidelines and standards. For instance, only aggregate statistical estimates that conform to the confidentiality provisions of the Statistics Act will be released outside of Statistics Canada.

Later releases will include an analytical report addressing important subjects like the use of aids and assistive devices, help received or required, and the employment and educational experiences of persons with disabilities.

In addition to these data products and services, clients can request custom data tables from Statistics Canada. All such requests are screened for confidentiality and the aggregate data are rounded before being released to clients. Statistics Canada also delivers special CSD presentations to key stakeholders and at various conferences.

9.2 Reference products

Information about the 2012 Canadian Survey on Disability (CSD) is available on Statistics Canada's website. Statistics Canada provides an Integrated Metadata Base (IMDB) on-line for all surveys that it conducts, including the 2012 CSD. The purpose of the IMDB is to provide information that will assist the public in interpreting Statistics Canada's published data. The information (also known as metadata) is provided to ensure an understanding of the basic concepts that define the data, including variables and classifications, the underlying statistical methods and surveys, and key aspects of the data quality. Direct access to the CSD questionnaire is also provided.

In addition to the IMDB, the present Concepts and Methods Guide is provided online for a detailed discussion of survey content, sampling design, data collection and processing, weighting of the data, data quality, differences between the 2012 CSD and the 2006 Participation and Activity Limitation Survey (PALS), and dissemination products for the CSD.

For researchers using the analytical files in Statistics Canada's Research Data Centres (RDCs), an RDC User Guide is available with detailed step-by-step instructions for using the data file. The RDC User Guide describes the structure of the data files in detail, including all core variables, derived variables and linkages to the National Household Survey (NHS). Detailed data dictionaries provide information for all variables available. The RDC User Guide also provides detailed guidelines for tabulation and statistical analysis, how to apply the necessary weights to the data, information on software packages available and guidelines for the release of data, such as rounding rules. The process of estimating the reliability of estimates, both quantitative and qualitative, is covered in detail.

9.3 Disclosure control

Statistics Canada is prohibited by law from releasing any data that would divulge information obtained under the *Statistics Act* that relates to any identifiable person, business or organization without the prior knowledge or the consent in writing of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data.

Appendix A – Survey indicators by questionnaire module

The list below includes descriptions of questionnaire indicators as well as derived variables (DVs) which combine different questionnaire items.

For a complete list of variables, please refer to the 2012 CSD Codebook (data dictionary). For details on how to obtain the codebook, please contact Statistics Canada Client Services or call 1-800-263-1136.

Disability screening questions (DSQ)

The DSQ measure different disability types and severities, asking questions about how often respondents' daily activities are limited by long-term conditions, health problems and task-based difficulties.

Disability filter questions

There are six filter questions on the DSQ. The first five questions ask about the frequency of difficulties or conditions that have lasted or are expected to last six months or more. These filters address five "families" of disability type, as follows:

1. Sensory: difficulty seeing or hearing
2. Physical (mobility, flexibility, dexterity, pain): difficulty walking, using stairs, using hands or fingers, or doing other physical activities
3. Cognitive (learning, memory or developmental): difficulty learning, remembering or concentrating
4. Mental/psychological health: emotional, psychological or mental health conditions
5. Other long-term health problem or condition

The sixth filter question on the DSQ asks how often the difficulty limits daily activities (or, if more than one difficulty is identified, the question asks about the most limiting health problem).

Filter questions on the DSQ were developed to filter out persons who are highly unlikely to have a disability. In the context of the 2012 Canadian Survey on Disability (CSD), initial filtering was already conducted using activity limitation questions on the 2011 National Household Survey (NHS). Thus the DSQ filter questions, although included on the 2012 CSD, were not used there for the purposes of filtering.

Disability screener questions

Screener questions on the DSQ evaluate the presence and severity of 10 distinct types of disabilities related to health problems or conditions that have lasted or are expected to last for six months or more. Questions ask how often a difficulty limits daily activities for the following disability types:

1. Seeing
2. Hearing
3. Mobility
4. Flexibility
5. Dexterity
6. Pain
7. Learning
8. Developmental
9. Mental/psychological
10. Memory

Disability screening derived variables include disability status, vision limitation status, hearing limitation status, mobility limitation status, agility limitation status, flexibility limitation status, dexterity limitation status, learning limitation status, pain limitation status, developmental limitation status, mental/psychological limitation status, memory limitation status, unknown type of limitation, and disability type counter.

Related DVs include severity class for vision, severity class for hearing, severity class for mobility, severity class for flexibility, severity class for dexterity, severity class for pain, severity class for learning, severity class for mental/psychological, severity class for memory, severity class for developmental, severity class for unknown, global severity score, and global severity class.

Disability questions

Main cause

This module is used to identify the main medical condition which caused the most difficulty or limitation to daily activity for the respondent. Respondents were able to identify up to two medical conditions that contributed to difficulties or activity limitations. Age of onset of difficulties or activity limitations was also asked.

Additional variables include International Classification of Diseases (ICD) for main conditions.

Aids and assistive devices

For each activity limitation identified in the disability screening questions (DSQ), questions were asked about any aids, assistive devices and specialized equipment that the respondent used and needed. Respondents who reported a need for an aid, assistive device or other type of specialized equipment that they did not have were then asked reasons for the unmet need.

Questions were also asked about out-of-pocket or direct expenses related to the purchase, repair or maintenance of aids, assistive devices and specialized equipment used by the respondent in the 12 months prior to the survey.

Derived variables include use of aids for hearing, needs for hearing aids, use of aids to help with vision, needs for aids to help with vision, use of aids to help with mobility or agility, needs for aids to help with mobility or agility, use of aids to help with learning, needs for aids to help with learning, use of any other aids to help with another condition, needs for aids to help with another condition, use of aids to help with any condition(s), and needs for aids to help with any condition.

Medication use

This module asked about prescription medications that respondents took for their condition. Respondents are also asked about the impact of cost on their medication use.

Help received for everyday activities

This module asked about the type of help that respondents received with everyday activities due to their condition. Respondents were asked to include help received from family, friends, neighbours, and from organizations, whether paid or unpaid. Everyday activities include meal preparation, daily housework, heavy household chores, getting to appointments and running errands, taking care of personal finances, personal care, help with basic medical care at home, moving about inside the home, and child care.

Respondents who did not usually receive help were asked if they thought they needed help with each activity. Respondents who reported receiving help with one or more activities were asked additional questions about the frequency of help received and about the persons or organizations that helped them with their everyday activities.

Respondents were also asked about out-of-pocket or direct expenses for help received for everyday activities in the past 12 months.

Note that this section concerns only "human" help, and not technical assistance.

Derived variables include needs help preparing meals, needs help with housework, needs help with heavy household chores, needs help with getting to appointments, needs help with personal finances, needs help with personal care, needs help with basic medical care, needs help with moving around, needs help with childcare, and receives help.

Education

Recent school attendance

This module asked about recent school attendance, full-time or part-time status, and the type of institution attended. Respondents who were currently attending school or who attended school at any time between September 2011 and the date of interview were asked additional questions about requirements for modified building features, specialized transportation, assistive devices or support services. If a need was identified, respondents were asked if accommodations were made available to them.

Past school attendance

This module asked about previous school attendance, and whether a respondent had the condition while attending school. Respondents with a condition who attended school at any time between September 2007 and the date of interview were asked additional questions about requirements for modified building features, specialized transportation, assistive devices or support services. If a need was identified, respondents were asked if accommodations were made available to them.

Educational experiences

The intent of this module was to determine the impact of a respondent's condition on his/her educational experience and level of schooling. Respondents were asked about difficulties encountered while acquiring an education. Questions asked about the impact of having a condition on choice of courses or careers, type or location of school, additional expenses and length of time to complete studies. Questions were also asked about the perceived social impact of the condition in terms of exclusion and bullying.

Educational background

This module addressed the highest level of education completed, location of study, and major field of study of the highest certificate, diploma or degree completed.

Additional education variables include major field of study (CIP Canada 2011).

Employment

Labour force status

This first module of the employment section determined which of three employment classifications – employed, unemployed, not in the labour force (includes retired) – applied to the respondent. This determined the flow of questions for the respondent throughout the remainder of the section.

Respondents who had worked in the week prior to the interview were classified as being employed. If they were absent from work because of an illness, illness in the family, vacation, maternity/paternity leave, or other reasons but still occupied a job, they were classified as employed.

Respondents who indicated that they were absent from a job due to a temporary lay-off, seasonal lay-off, or injury or sick leave that was no longer covered by the employer, they were classified as being unemployed at the time of the interview. If a respondent had arrangements to start a job within four weeks of the time of interview, they were classified as unemployed.

Respondents were classified as not in the labour force if they did not have a job and were unable to start a job in the next four weeks due to school, illness/disability, personal and family responsibilities, or were not looking for work.

Persons who were retired were included under the classification of "not in the labour force" for the initial set of questions and were then identified as "retired" for future sections.

Employment details

Respondents were asked questions about their current job, including hours worked, employment status (employee or self-employed), details about their business (name of business, kind of business, industry, number of employees in the workplace, etc.) and occupation.

Respondents were also asked about the impact of their condition on their work: whether having a condition limited the amount or kind of work they were able to do, and whether their condition affected their ability to change jobs or advance at their present job.

The Unemployed

Questions in this module determined when a respondent last worked, and asked questions about previous employment, including hours worked, employment status (employee or self-employed), details about the business (name of business, kind of business, industry, etc) and occupation. Respondents were asked whether having a condition limited the amount or kind of work they were able to do, and whether their condition affected their ability to look for work.

Not in labour force

Questions in this module determined when a respondent last worked. A respondent was included in the "not in the labour force" category if he or she did not have a job and was unable to start a job within the next four weeks due to school, illness/disability, personal or family responsibilities or was not looking for work.

Respondents who had never worked were asked follow-up questions about whether their condition prevented them from working. Respondents who were permanently retired were asked if this was due to their condition.

Respondents who reported working in the past were asked questions about their previous employment, including hours worked, employment status (employee or self-employed), details about their business (name of business, kind of business, industry, etc) and occupation.

Respondents were asked whether having a condition prevented them from working, or limited the amount or kind of work they were able to do. Respondents were also asked about barriers to looking for work.

It is in this module that we distinguished between people who were not in the labour force and those who were retired.

Retirement

Respondents were asked the year in which they first retired, whether retirement was voluntary, and whether having a condition prevented them from working, or limited the amount or kind of work they were able to do. Respondents who report a non-voluntary retirement were asked a follow-up question about barriers to looking for work.

Workplace training

The questions in this module dealt with job-related training provided or paid for by a respondent's employer or company in the 12 months prior to the survey. A question about barriers to taking work-related training was also asked.

Job-related training included activities that were pursued for the development or upgrading of skills for use in present or future employment rather than personal interest or other non-employment-related reasons. Training activities included courses and workshops not leading to a formal education credential, or activities (programs) provided in formal systems of education leading to a formal education credential, including primary-level and secondary-level education, and postsecondary education like university and college diplomas and degrees.

Employment modifications

This module asks about whether a respondent required, or would require modifications or changes to his or her job or physical work environment to be able to work. Follow-up questions asked about whether the modification or change had been made available, and if not, reasons why accommodations had not been provided.

Labour force discrimination

These questions asked about the respondent's perception of disadvantage in the work force, for example, being refused job interviews, jobs or promotions.

Employment DVs include labour force status, type of industry of person's work (North American Industry Classification System (NAICS) Canada 2007), type of occupation (National Occupational Classification (NOC) 2011), full-time or part-time employment status, year of tenure at current job or business, year when last worked, and year of first retirement.

Getting around the community

This module asked about local travel using public and specialized transit services. The objective of this module was to identify any difficulties the respondent had in using these methods of transportation, for example, because it was not available or not accessible.

Source of income

This module asked about personal income received from various sources including wages and salaries, disability benefits, employment insurance and social assistance.

A DV for source of income summarizes employment and non-employment sources.

Appendix B – Disability Severity Score

Severity score

A severity score was developed using the Disability Screening Questions (DSQ). The method used to derive a score for each disability type and an overall score that takes all disability types into account is described below.

The score had to satisfy the following three criteria:

- it had to increase with the number of disability types;
- it had to increase with the level of difficulty associated with the disability;
- it had to increase with the frequency of the activity limitation.

Score for each disability type

Using a scoring grid that takes into account both the frequency of the activity limitations (never, rarely, sometimes, often, or always) and the intensity of the difficulties (no difficulty, some difficulty, a lot of difficulty, or cannot do), a score is assigned to each of the 10 disability types. The score increases with the frequency of the limitation and the level of difficulty. Hence, a person who reports always being limited in his or her activities and being unable to do an activity will have the maximum score for that disability type. The scoring grid developed for each disability type is shown below. In the grid, the most severe responses have higher scores, moderate responses have mid-range scores, and mild responses have lower scores.

Table B.1
Scoring grid based on intensity of difficulty and frequency of daily activity limitations

Intensity of difficulty	Frequency of daily activity limitations				
	Never	Rarely	Sometimes	Often	Always
No difficulty	0	0	2 ¹	3 ¹	4 ¹
Some difficulty	0	0	4 ²	6 ²	12 ³
A lot of difficulty	0	3 ¹	6 ²	18 ³	20 ³
Cannot do	0	4 ²	12 ³	20 ³	24 ³

1. Mild responses

2. Moderate responses

3. Severe responses

Source: Statistics Canada, Canadian Survey on Disability, 2012.

The score for each disability type is then normalized to a value between 0 and 1. This is done by dividing the score in each cell by the maximum score of 24. The table below shows the normalized score for each disability type.

Table B.2

Normalized score based on intensity of difficulty and frequency of daily activity limitations

Intensity of difficulty	Frequency of daily activity limitations				
	Never	Rarely	Sometimes	Often	Always
No difficulty	0	0	0.0833 ¹	0.1250 ¹	0.1667 ²
Some difficulty	0	0	0.1667 ²	0.2500 ³	0.5000 ³
A lot of difficulty	0	0.1250 ¹	0.2500 ²	0.7500 ³	0.8333 ³
Cannot do	0	0.1667 ²	0.5000 ³	0.8333 ³	1.0000 ³

1. Mild responses

2. Moderate responses

3. Severe responses

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Special cases

The scoring grid shown above was used for the majority of the disability types, but it had to be adapted for three special situations.

First, the DSQ had no questions on the intensity of the difficulty for the "unknown" type. Consequently, only those who reported being limited "sometimes", "often" or "always" were considered to have a disability. Since, by definition, people in the "unknown" type have only one type of disability, we examined the intensity distribution for other people who had only one of the 10 disability types. We found that in most cases, the intensity level was "some difficulty". For the "unknown" type, therefore, we used the scoring grid shown above and set the intensity level at "some difficulty". The "sometimes" responses were hence given 0.1667 points, the "often" responses 0.25 points, and the "always" responses 0.50 points.

The second special case concerns developmental disorders. For this disability type, a person who reports being diagnosed with this disorder is automatically deemed to have a disability, regardless of the level of difficulty and the frequency of the limitations reported. However, in calculating a score for this type, we want to avoid situations where people with a developmental disorder have a score that is very low or 0. We therefore set a minimum score for people with developmental disorders. We use the grid to determine the score, but if the reported frequency of the limitation is "never", or if it is "rarely" combined with "no difficulty" or "some difficulty", we set the score at 0.0833, which is the minimum score that a person can have for any other disability type. We assigned a minimum score in about one of every five cases with developmental disorders.

The third special case relates to the "mobility" and "flexibility" types. For these two types, we have two questions on the intensity of the difficulty, instead of just one question. To calculate the score for these two types, we used the higher intensity reported on the two questions, combined with the frequency of the limitation, to determine the score. The decision was made to use the higher intensity for the two questions, rather than an average value, for example, because we wanted each of the 10 disability types to have the same importance in the calculation of the overall score. The fact that we needed two questions to establish whether the respondent has difficulty for mobility or for flexibility should not lower the person's score when the answer to one question is "no difficulty" and the answer to the other is "cannot do", for example.

Severity class for each disability type

For **each disability type**, we define two severity classes: less severe and more severe. People with a score of 0.5 or more for a disability type are assigned to the more severe class, and others are assigned to the less severe class. People with a score of 0 for a disability type are classified as not having that disability type.

We created a severity class variable for each disability type (**DVIS_CL** for seeing, **DHEA_CL** for hearing, **DMOB_CL** for mobility, **DFLE_CL** for flexibility, **DDEX_CL** for dexterity, **DPAL_CL** for pain, **DLEA_CL** for learning, **DDEV_CL** for developmental, **DEMO_CL** for emotional, **DMEM_CL** for memory, and **DUNK_CL** for unknown). Each variable takes one of the following values:

- 0 = no disability
- 1 = less severe
- 2 = more severe

Overall score

A person's **overall severity score** is calculated by taking the average of the scores for the 10 disability types. Consequently, the more types of disability a person has, the higher his or her score will be. By definition, the overall score is also a number between 0 and 1. People who report no disability for the 10 types but report another health problem or condition (i.e., they have an "unknown" type) have an overall score equal to the score for the "unknown" type divided by 10.

We created a variable for the overall score: **DScore**.

Overall severity class

To make the severity score easier to use, severity classes were established. It is important to understand that the name assigned to each class is simply intended to facilitate its use. It is not a label or judgement concerning the person's level of disability. In other words, the classes should be interpreted as follows: people in class 1 have a less severe disability than people in class 2; the latter have a less severe disability than people in class 3; and so on. We created a variable, **DCLASS**, for the overall severity class:

- 1 = mild disability
- 2 = moderate disability
- 3 = severe disability
- 4 = very severe disability

The cut-off point between classes 2 and 3 was first determined graphically. There is a natural discontinuity in the distribution of the overall score near the value of 0.1. In fact, this value corresponds to the score of a person who received the maximum number of points for a single disability type. That would put the person in the severe class. To be classified as very severe, a person has to have double that score, or 0.2. A score of 0.2 corresponds to the score of a person who received the maximum number of points for two disability types. The cut-off point between class 1 and class 2 is half the score required for class 3, i.e., 0.05. That corresponds to the score of a person who is in one of the most severe cells for one disability type, for example.

The breakdown of persons with a disability across the four severity classes is shown in the table below.

Table B.3
Distribution of persons with a disability, by severity class, Canada

Severity class	Persons with a disability	
	number	%
Class 1 = mild	1,195,590	31.7
Class 2 = moderate	747,980	19.8
Class 3 = severe	849,540	22.5
Class 4 = very severe	982,810	26.0

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Overall score for persons without a disability in the CSD

In the CSD's sample (the YES sample), there are two types of people without a disability:

- People interviewed for the CSD who reported being limited only "rarely" with "no difficulty" or "some difficulty". For the purposes of the CSD interview, these people were deemed potentially to have a disability and therefore had to answer all the survey questions. However, they were excluded from the final definition of disability.
- People interviewed for the CSD who reported that they were "never" limited (false positives). For the purposes of the CSD interview, these people were deemed not to have a disability and therefore did not have to answer the rest of the survey questions.

We calculated an overall score for the people without a disability in group (a), but the people in group (b) all have a score of 0. Since CSD data are available for the people in group (a), users may want to study their characteristics and try to compare the severity of the group's members among themselves, even though all these cases are very mild. The number of points used to calculate the scores of these persons was intentionally set very low to ensure that people who reported being "rarely" limited with "no difficulty" or "some difficulty" for several types would not end up with a higher score than someone who is considered to have a disability under the definition used in the CSD.

Table B.4
Scores used for persons without a disability from group (a)

Intensity of difficulty	Frequency of daily activity limitations				
	Never	Rarely	Sometimes	Often	Always
No difficulty	-	0.1	-	-	-
Some difficulty	-	0.2	-	-	-
A lot of difficulty	-	-	-	-	-
Cannot do	-	-	-	-	-

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Again, the score for each disability type is then normalized to a value between 0 and 1 by dividing it by the maximum score of 24. The table below shows the normalized scores of people who are considered not to have a disability but answered all of the CSD question.

Table B.5

Normalized scores used for persons without a disability from group (a)

Intensity of difficulty	Frequency of daily activity limitations				
	Never	Rarely	Sometimes	Often	Always
No difficulty	-	0.0042	-	-	-
Some difficulty	-	0.0083	-	-	-
A lot of difficulty	-	-	-	-	-
Cannot do	-	-	-	-	-

Source: Statistics Canada, Canadian Survey on Disability, 2012.

It is important to note, however, that for people without a disability in group (a), even though we had to calculate a score for each disability type to obtain the overall score, the score for each type was then reset to 0 in the final file. The number of disability types was also set to 0. This is to avoid any confusion when the types of disability are studied, since these people should not be counted as having a disability.

Appendix C – Employment equity questions

Persons who have a long-term disability

Filter questions: DSQ_Q01-Q06

Disadvantage in Employment

EDE_Q15, UDE_Q12, NDE_Q13: Does your condition limit the amount or kind of work you can do at your present job or business?

EDI_Q04: Do you consider yourself to be disadvantaged in employment because of your condition?

Perception

EDE_Q19: Do you believe that your condition makes it difficult for you to change jobs or to advance at your present job?

EDI_Q05: Do you believe that your current employer or any potential employer would be likely to consider you disadvantaged in employment because of your condition?

Accommodation

EMO_Q01: Because of your condition, do you require any of the following to be able to work?

1. Job redesign (modified or different duties)
2. Telework
3. Modified hours or days or reduced work hours
4. Human support, such as a reader, Sign language interpreter, job coach or personal assistant
5. Technical aids, such as a voice synthesizer, a TTY, an infrared system or portable note-taker
6. A computer or laptop with specialized software or other adaptations such as Braille, screen magnification software, voice recognition software or a scanner
7. Communication aids, such as Braille or large print reading material or recording equipment
8. A modified or ergonomic workstation
9. A special chair/ back support
10. Handrails, ramps or widened doorways or hallways
11. Adapted or accessible parking
12. Accessible elevator
13. Adapted washrooms
14. Specialized transportation
15. Other equipment, help or work arrangement

Appendix D – Extra coding categories and standard classifications used

Extra response categories created for “Other – specify” questions

The table below summarizes the extra categories added to certain survey questions during the coding of data from questions involving an “Other – specify” category. Where counts warranted, new categories were created to capture themes in the data that were not reflected in existing categories. These categories are thus additions to those already appearing on the questionnaire of the Canadian Survey on Disability and will be available on the survey database.

Table D.1
Extra response categories created for “Other – specify” questions

Theme	Indicator	Questions	Extra categories
Aids and assistive devices	Aids or assistive devices needed but does not have	AAD_Q10	-Glasses and/or contact lenses
Aids and assistive devices	Aids, devices or specialized equipment needed but does not have	AAD_Q27	-Aids for persons who are deaf or hard of hearing -Mobility Aids such as wheelchairs, scooters, canes and walkers

Source: Statistics Canada, Canadian Survey on Disability, 2012.

Standard classifications used for the 2012 CSD

Listed below are the standard classifications used in the coding of open-ended questions on the Canadian Survey on Disability (CSD).

Medical condition: International Classification of Diseases, Tenth Revision (ICD-10)

CSD respondents provided information on the main medical condition which caused them the most difficulty or limitations to their daily activities. Responses were coded to the International Classification of Diseases, Tenth Revision (ICD-10).

The ICD is the standard diagnostic tool for epidemiology, health management and clinical purposes. It allows for the analysis of the general health situation of population groups. The ICD is used to monitor the incidence and prevalence of diseases and other health problems.

ICD-10 was endorsed by the 43rd World Health Assembly in May 1990 and came into use in World Health Organization (WHO) Member States as from 1994.

Major field of study: Classification of Instructional Programs (CIP)

Respondents to the 2012 CSD provided information on the major field of study of the highest certificate, diploma or degree that they completed. Responses were coded to the Classification of Instructional Programs (CIP) Canada 2011.

The CIP was originally developed in 1980 by the National Center for Education Statistics (NCES) in the United States. NCES released updates in 1985, 1990, 2000 and 2010. CIP Canada 2011 is the second Canadian version of this classification; the first Canadian version being CIP Canada 2000.

CIP is designed to classify ‘instructional programs’. The organizing principle behind CIP is ‘field of study’. At Statistics Canada, a field of study is defined as a “discipline or area of learning or training”.

Industry and occupation: NAICS and NOCS

CSD respondents provided information on their occupation and the industry in which they worked during the reference week of the survey. These responses were coded using the North American Industry Classification System (NAICS) Canada and the National Occupational Classification (NOC). Data were coded to the 4-digit level of detail and were coded to both current/emerging and historical versions of these classification systems, as follows:

NAICS Canada 2012 - North American Industry Classification System (NAICS) Canada

NAICS Canada 2007 - North American Industry Classification System (NAICS) Canada

NOC – 2011: National Occupational Classification (NOC)

NOC-S – 2006: National Occupational Classification - Statistics (NOC-S)

The North American Industry Classification System (NAICS) is an industry classification system developed by the statistical agencies of Canada, Mexico and the United States. It was created against the background of the North American Free Trade Agreement. The 2012 NAICS revision was undertaken to achieve one main goal: to modify or create industries to reflect new, emerging, or changing activities and technologies. NAICS is based on supply-side or production-oriented principles - the criteria used to group establishments into industries in NAICS are similarity of input structures, labour skills and production processes.

NOC 2011 is the nationally accepted taxonomy and organizational framework of occupations in the Canadian labour market. It updates both the National Occupational Classification of Employment and Social Development Canada and Statistics Canada's National Occupational Classification for Statistics (NOC-S). As in the past, this revision is based on extensive occupational research, analysis and consultation conducted across the country. Occupations are identified and grouped primarily in terms of the work usually performed, this being determined by the tasks, duties, and responsibilities of the occupation.

Appendix E – Glossary of survey terms

A

Analytical file

A Statistics Canada **microdata** set for a given survey, available for use in Research Data Centres (RDCs) across Canada. RDCs provide researchers with access, in a secure university setting, to microdata from population and household surveys. The centres are staffed by Statistics Canada employees. They are operated under the provisions of the **Statistics Act** in accordance with all the confidentiality rules and are accessible only to researchers with approved projects who have been sworn in under the Statistics Act as 'deemed employees.'

B

Bootstrap method

The bootstrap method is an approach for estimating error in a dataset related to **sampling**. Sampling introduces error because data are not taken from the entire population, but only a sub-section, called a sample, which is then used to make **estimates** for the whole population. There are several methods for estimating the level of **sampling error**. The bootstrap method usually selects a number of subsamples from the main sample and produces estimates for each subsample. The sampling error is estimated as a function of the observed differences between estimates from the different subsamples and estimates from the complete sample.

C

Census metropolitan area (CMA) and Census agglomeration (CA)

Area consisting of one or more neighbouring municipalities situated around a major urban core. A census metropolitan area must have a total population of at least 100,000 of which 50,000 or more live in the urban core. A census agglomeration must have an urban core population of at least 10,000.

Census of population

A census is the collection of information about all units in a population, sometimes also called a 100% sample survey. Under the **Statistics Act** of 1971, it is a statutory requirement to conduct a nationwide census every five years. The Census of Population provides information needed to plan community services such as schools, day care, police services and fire protection, to forecast consumer demand and to conduct market research studies.

Cohort

As used in demography, a number of people having a common characteristic, for example, all persons in a given population who were born in 1940, or all persons suffering from a particular disease.

Confidential information

This is a term used within Statistics Canada to describe information that is subject to the secrecy provisions of the **Statistics Act**. Information is deemed confidential either because it directly identifies a **responding unit**, for example, by name, or because it could permit specific responding units to be identified, even when the data is stripped of identifiers, due to the information's detail or its geographical structure or format.

Confidentiality

Confidentiality denotes an implied trust relationship between the person providing the information and the individual or organization collecting it. This relationship is built on the assurance that the information will not be disclosed without the person's permission. Under the **Statistics Act**, information that would identify an individual, business or institution cannot be disclosed without their knowledge or consent.

Coverage

Coverage is the extent to which every person or unit intended for inclusion in a survey or census is in fact counted and counted only once. Coverage errors refer to when persons or units of the survey or census are missed (under-coverage) or over-counted (over-coverage). Studies are often conducted by Statistics Canada to provide estimates of under-coverage and over-coverage of a given survey or census or to examine related issues. For example, Statistics Canada has studied and analyzed the extent to which cell-phone use affects coverage for telephone surveys.

CV – Coefficient of variation

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. In this process of estimation, some level of error is inevitable. The coefficient of variation (CV) is a way of expressing the **sampling error** associated with an **estimate**. First a **standard error** or 'average' error of the estimate is calculated. The CV is obtained by dividing the standard error of the estimate by the estimate itself and expressing the resulting fraction as a percentage. The lower the CV, the higher the data quality (see **margin of error**).

D

Data

Observations and measurements collected during a survey, census or other study. Facts or figures from which conclusions can be drawn.

Data quality

A degree or level of confidence that the data and statistical information are "fit for use". The particular issues of quality or fitness for use that must be addressed by Statistics Canada are relevance, accuracy, timeliness, accessibility, **interpretability** and coherence.

Dataset, database

An organized and sorted list of facts or information about a set of individuals, households, businesses, or other relevant units. A Statistics Canada dataset is usually generated by a survey or administrative data, stored on a computer, and organized in such a way that it may be accessed easily by a wide variety of statistical application programs.

Dissemination

The process of providing statistical products and services to the general public and to specific data users. Statistics Canada disseminates data and analysis in the form of survey results, research reports, technical papers, periodical magazines, census products, and research compendia. Online products date from 1996 to the present. Historical material can be located using the Library Catalogue. Statistics Canada information is also distributed to an approved network of depository libraries.

The objective of dissemination activities is to provide relevant information in a timely fashion, in useful formats, and through accessible channels. Activities in place to support the dissemination of products include client consultation services, marketing, promotions, user-training and other client services.

Derived variable

A new **variable** constructed by applying logical or mathematical operations to one or more existing variables in order to meet particular data needs. For example, an age variable can be derived from date of birth information. As another example, a derived variable could be obtained called 'presence of a chronic health condition' based

on whether or not a respondent answered 'yes' at least once to a series of questions asking about specific chronic health conditions such as asthma, diabetes, heart disease, etc.

E**Editing**

Editing is a process that ensures survey data are accurate, complete and consistent. A set of editing rules or conditions is applied to a **dataset**. Data which do not meet the conditions are examined and corrected where appropriate.

Errors

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. The accuracy of such an **estimate** is a measure of how much the estimate differs from the correct or "true" figure. Departures from true figures are known as errors. Errors can arise from many sources, but can be grouped into a few broad categories: coverage errors, non-response errors, response errors, processing errors and sampling errors.

Coverage errors

Coverage errors refer to when persons or units of the survey are missed (under-coverage) or overcounted (over-coverage).

Non-response errors

Non-response errors occur when it proves impossible to obtain a complete questionnaire from a person, household, or organization. Although certain adjustments for missing data can be made during processing, non-response means some loss of accuracy is inevitable.

Response errors

Response errors indicate that a response may not be entirely accurate. The respondent may have misinterpreted the question or may not know the answer, especially if it is given for an absent household member, for example.

Processing errors

Processing errors include mistakes made during data entry, coding, tabulation or other forms of data manipulation.

Sampling error

Sampling error refers to the fact that the results of the weighted sample differ somewhat from the results that would have been obtained from the total population. The difference is known as sampling error. The actual sampling error is of course unknown, but it is possible to calculate an "average" value, known as the "**standard error**".

Estimation, estimate

Using results of the weighted sample to estimate the characteristics of the total population.

F

Frame

A list, map, or conceptual specification of the units comprising the survey population from which persons can be selected. For example, a telephone or city directory, or a list of members of a particular association or group.

Frequency

The number of times an event or item occurs in a **dataset**.

Frequency distribution

A chart or table showing how often each value or range of values of a **variable** appear in a **dataset**. It is sometimes called a one-way frequency table to indicate that the distribution contains counts for one variable only.

G - H - I

Imputation

Imputation involves replacing either missing or invalid data with valid data. This is normally performed using predetermined rules or with the use of data from a 'statistical neighbour'—another **responding unit** who has similar characteristics. Imputation is often combined with data **editing**.

In scope

A **unit** that meets all criteria for the survey

Information

Data that have been recorded, classified, organized, related or interpreted within a framework so that meaning emerges.

Information product

Results of Statistics Canada activities that are organized into data files, **databases**, tables, graphs, maps, and text. This organization can be either pre-defined (standard information product) or made in response to special requests (customized information product). Information products can be made available on either print or electronic media.

Interpretability

Interpretability reflects the ease with which the user may understand, properly use and analyze the data or information. The degree of interpretability is largely determined by: the adequacy of definitions on concepts, target populations and variables; terminology underlying the data; and information on any limitations of the data.

Institution

Institutional collective dwellings are general and speciality hospitals, chronic care and long-term care hospitals, nursing homes, group homes or institutions for the physically disabled and treatment centres, group homes for children and youth, group homes or institutions for people with psychiatric disorders or developmental disabilities, federal correctional institutions, provincial and territorial custodial facilities, young offenders' facilities, jails and police lock-up facilities, shelters for persons lacking a fixed address, shelters for abused women and their children and other shelters and lodging with assistance services.

J - K - L

Logistic regression

A form of **regression** analysis used when the response variable is a binary **variable** (a variable having two possible values).

M

Margin of error

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. In this process of **estimation**, some level of error is inevitable. The margin of error, a measure used to build confidence intervals, serves as a rough indicator of the precision of an estimate.

For example, pollsters often say that a certain percentage of the population, plus or minus the margin of error (expressed in percentage points), is likely to vote for a certain candidate, 19 times out of 20. To calculate the margin of error, which in this example corresponds to a 95% confidence interval, the pollster would use the equivalent of plus or minus two standard errors of the estimate (see **Standard error**).

Methodology

A set of research methods and techniques applied to a particular field of study. At Statistics Canada, methodology refers to survey methodology.

Microdata

Files of **records** pertaining to individual **responding** units.

N - O

National Household Survey

The National Household Survey (NHS) was introduced in 2011 as a replacement for the long census questionnaire, more widely known as Census Form 2B/2D. The NHS is designed to collect social and economic data about the Canadian population. The objective of the NHS is to provide data for small geographic areas and small population groups.

Observation

Data collected for a given variable about a particular **responding unit**. Examples include the specific values for a responding unit on characteristics such as age, gender or marital status—the observation might be '77', 'woman' and 'widowed'.

Out of scope

A **sampled unit** that does not meet all criteria for being surveyed.

P

Population

The complete group of units to which survey results are to apply. These units may be persons, households, businesses, institutions, etc. The term "Target Population" is often used to refer to all potentially **surveyed units**, as defined in a clear, precise way by the survey study. This is the population for which information is wanted.

Population Centre : The term population centre replaces the term urban area (as used in the Census of Population until 2006). A population centre is defined as an area with a population of at least 1,000 and no fewer than 400 persons per square kilometre. Population centres are classified into three groups, depending on the size of their population:

- small population centres, with a population between 1,000 and 29,999
- medium population centres, with a population between 30,000 and 99,999
- large urban population centres, with a population of 100,000 or more.

Postcensal survey

A postcensal survey is one where **surveyed units** are selected based upon their responses to the **Census of Population or the National Household Survey (NHS)**. These surveys are generally conducted shortly after the Census or NHS.

Proportion

A proportion refers to how many responses fall into a given response category in relation to the total responses. It is calculated by dividing the frequency of the response category by the total number of responses to the question.

Q – R

Record

A record is the data for an individual **responding unit** in a file containing data for all of a survey's responding units.

Regression

A statistical method which tries to predict the value of a characteristic by studying its relationship with one or more other characteristics. This relationship is expressed through the means of a regression equation.

Research data centres

Research data centres (RDCs) provide researchers with access, in a secure Statistics Canada governed setting, to micro data from population and household surveys. The RDC program is part of an initiative by Statistics Canada, the Social Sciences and Humanities Research Council (SSHRC) and university consortia to help strengthen Canada's social research capacity and to support the policy research community. The program is also supported by the Canadian Foundation for Innovation (CFI) and the Canadian Institute of Health Research (CIHR).

Respondent

The respondent is the person providing the information for the **surveyed unit**, which could be a person, household, business or institution

Response rate

The proportion of a sample for which a response to a questionnaire is obtained, usually expressed as a percentage. Non-response covers those who refused to participate as well as persons whom the survey was unable to reach.

Sample design

A set of specifications that describe the sampling elements of a survey in detail. These elements include **population, frame, surveyed units, sample size, sample selection and estimation method**.

Sampling

The process of selecting some part of a population to observe so as to estimate something of interest about the whole population. Examples of different sampling methods include simple random sampling, stratified random sampling, cluster sampling and multi-stage sampling.

Sampling rate/Sampling fraction

Sample size divided by the population size.

Sampling or sampled unit

The unit selected by the **sample design** and from which measurements are taken for a survey. Examples include persons, households, families or businesses. For the CSD, the sampling unit is the person.

Standard deviation

Standard deviation measures the dispersion of a data set around the mean. It is the most widely-used measure of dispersion. Mathematically, the standard deviation is the square root of variance.

Standard error

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. **Sampling error** refers to the fact that the results of the weighted sample differ somewhat from the results that would have been obtained from the total population. The difference is known as sampling error. The actual sampling error is of course unknown, but it is possible to calculate an "average" value, known as the "standard error".

Statistics Act

An Act regarding statistics of Canada. It includes the definition of Statistics Canada's mandate: "There shall continue to be a statistics bureau under the Minister, to be known as Statistics Canada, the duties of which are:

- to collect, compile, analyze, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people;
- to collaborate with departments of government in the collection, compilation and publication of statistical information, including statistics derived from the activities of those departments;
- to take the census of population of Canada and the census of agriculture of Canada as provided in this Act;
- to promote the avoidance of duplication in the information collected by departments of government; and,
- generally, to promote and develop integrated social and economic statistics pertaining to the whole of Canada and to each of the provinces thereof and to coordinate plans for the integration of those statistics."

Stratified sampling, stratification

A sampling procedure in which the population is divided into homogeneous subgroups or strata and the selection of samples is done independently in each stratum.

Suppression

The process by which particular data are prevented from being released based on criteria designed to protect confidentiality. 'Cell' suppression refers to procedures used to protect sensitive tabular data from disclosure; a cell being an individual entry in a table. For the CSD, data were also suppressed for reasons of data quality (CV larger than 33.3%).

Surveyed unit

The selected unit from which measurements are taken for a sample survey or a Census. Examples include persons, households, families or businesses. For the CSD, the surveyed unit (which is also the sampled unit since the CSD is a sample survey) is a person.

T - U - V

Unit

Same as **surveyed unit**.

User guides

These guides accompany Statistics Canada survey datasets, such as the **Analytical Files**, providing the detailed technical information required to use the data appropriately. The guide typically contains important information to know prior to data analysis: weighting variables to use, procedures related to the estimate of variance, and precautions to take in the dissemination of the data.

Valid skip

Indicates that a survey question was skipped because it did not apply to the respondent's situation, as determined by valid answers to a previous question. In such cases, the respondent is not considered to be part of the target population or universe for that question.

Variable

A characteristic that may assume more than one value to which a numerical measure can be assigned (e.g., income, age and weight).

Variance

A measure of dispersion for a given characteristic or variable in a dataset. It indicates how much variability exists for that characteristic. Technically, it is calculated as the average squared deviation from the mean of each observation in the data set for a particular variable.

References

- Bernier, J., and Nobrega, K. (1998). "Outlier detection in asymmetric samples: A comparison of an inter-quartile range method and a variation of a sigma-gap method". Annual meeting of the Statistical Society of Canada, June 1998.
- Federal Disability Report: The Government of Canada's Annual Report on Disability Issues (2010). Human Resources and Skills Development Canada, Gatineau, Québec.
- Haddou, M. (2013). "Bootstrap Variance Estimation Specifications - Aboriginal Peoples Survey". Internal document, January 2013.
- Langlet, É., Beaumont, J.-F., and Lavalée, P. (2008). "Bootstrap Methods for Two-Phase Sampling Applicable to Postcensal Surveys". Paper presented at Statistics Canada's Advisory Committee on Statistical Methods, May 2008, Ottawa.
- Mackenzie, Andrew, Matt Hurst and Susan Crompton (2009). Living with disability series, *Defining disability in the Participation and Activity Limitation Survey*, Canadian Social Trends, 2009/12. Statistics Canada catalogue no. 11-008-X.
- Morel, J., and Nambu, C. (2013). "National Household Survey: Weighting and estimation update". Presented at the SSC conference in May 2013.